

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:45 ; Search time 49.92 Seconds  
(without alignments)  
225.356 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1676  
Sequence: 1 EDCELPRLRRMTETILGSWS.....PDIKHGGLYHNMRRPYFPV 329

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A.Geneseq.36.\*  
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1362	72.6	240	18	Human partial Comp
2	1229	65.5	216	18	Clone PRB9PH410 C
3	353	18.8	496	20	Human CR1 protein
4	353	18.8	581	12	Human C4 binding p
5	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	350.5	18.7	778	19	Amino acid sequenc
7	350.5	18.7	1930	19	Human complement r
8	350.5	18.7	2039	20	Human C3b/C4b rece
9	350.5	18.7	2317	10	CR1 protein. Homo
10	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	343.5	18.3	543	13	CR1-4 (114S) analo

13	341.5	18.2	543	13	R28550
14	341.5	18.2	543	13	R28553
15	341.5	18.2	543	13	R28558
16	341.5	18.2	543	13	R28568
17	341.5	18.2	543	13	R28569
18	341.5	18.2	543	13	R28571
19	341.5	18.2	2039	12	R11810
20	340.5	18.2	543	13	R28545
21	340.5	18.2	543	13	R28548
22	339.5	18.1	543	13	R28555
23	339.5	18.1	543	13	R28558
24	339.5	18.1	543	13	R28567
25	339.5	18.1	543	13	R28563
26	338.5	18.0	543	13	R28549
27	338.5	18.0	543	13	R28551
28	338.5	18.0	543	13	R28562
29	336.5	17.9	543	13	R28556
30	336.5	17.9	543	13	R28564
31	336.5	17.9	2039	14	R36743
32	336	17.9	1537	12	R11982
33	335.5	17.9	543	13	R28552
34	335.5	17.9	543	13	R28561
35	335.5	17.9	543	13	R28566
36	334.5	17.8	543	13	R28554
37	333.5	17.8	543	13	R28544
38	333.5	17.8	543	13	R28546
39	332.5	17.7	543	13	R28543
40	332.5	17.7	543	13	R28559
41	330	17.6	646	20	R28558
42	329	17.5	453	20	R28557
43	327.5	17.5	450	20	R28555
44	327	17.4	450	20	R28554
45	317	16.9	577	17	R06882

#### ALIGNMENTS

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p
5	W39154	standard; Protein; 240 AA.	352.5	18.8	543	13	CR1-4 (266-274 KLR
6	W39154	standard; Protein; 240 AA.	350.5	18.7	778	19	Amino acid sequenc
7	W39154	standard; Protein; 240 AA.	350.5	18.7	1930	19	Human complement r
8	W39154	standard; Protein; 240 AA.	350.5	18.7	2039	20	Human C3b/C4b rece
9	W39154	standard; Protein; 240 AA.	350.5	18.7	2317	10	CR1 protein. Homo
10	W39154	standard; Protein; 240 AA.	347.5	18.5	543	13	CR1-4 (99H, 103E)
11	W39154	standard; Protein; 240 AA.	345.5	18.4	543	13	CR1-4 (52S, 53S, 5
12	W39154	standard; Protein; 240 AA.	343.5	18.3	543	13	CR1-4 (114S) analo

RESULT	ID	Description	Score	Query Match	Length	ID	Description
1	W39154	standard; Protein; 240 AA.	1362	72.6	240	18	Human partial Comp
2	W39154	standard; Protein; 240 AA.	1229	65.5	216	18	Clone PRB9PH410 C
3	W39154	standard; Protein; 240 AA.	353	18.8	496	20	Human CR1 protein
4	W39154	standard; Protein; 240 AA.	353	18.8	581	12	Human C4 binding p

PT H related antigen, or nucleic acid encoding it  
 XX  
 XX  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone PRB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigens can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.  
 CC  
 SQ Sequence 240 AA.

Query Match 72.6%; Score 1362; DB 18; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-96;  
 Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 78 FFLTGNVEYGVKAVYTCNEGQLLGETINRECDTGDWNTDIPICEVVKCLPVTAPENG 137  
 DB 1 flltgnvteygvkavylcneqyqlldetnrecdtdgwtndipicevkcldpvtapeng 60  
 QY 138 KIVSSAMEPDREYHFGQAVRFVCSNGYKTEGDEMHCSDDGFWSKPKVCVEISCKSPDV 197  
 DB 61 kivsamedpreyhfgavrfvcsngyklegdeemhcsddgfwskpkvcveisckspdv 120  
 QY 198 INGSPISSOKIITKENDREFOYKCNMGYEYSEKDAVCTESGMRPLPSCCEKSCDNPYIPNG 257  
 DB 121 inspsisqkilykenerfykcnmgysersgavctesgmrplpsceekscdnpyipng 180  
 QY 258 DYSPIRIKHTDELTITOCORNGEYPATRGNTAKCTSTGIPAPRCTLKACDIPDITKAGCL 317  
 DB 181 dyspirikhtgdeltytorgngfypatrgntakctstgwaprccltkpcdypdikhagl 240

## RESULT 2

W39155  
 ID W39155 standard; Protein: 216 AA.

AC W39155;

DT 27-APR-1998 (first entry)

DE Clone PRB9FH410 CFH related protein fragment.

KW Complement factor H; tumour associated antigen; renal cancer;  
 KW urogenital cancer; medicament; modulator.

OS Synthetic.

PN W09738136-A1.

PD 16-OCT-1997.

PE 09-APR-1997; 97WO-US05710.

PR 06-MAR-1997; 97US-0812481.

PR 09-APR-1996; 96US-0015083.

PR 06-MAR-1997; 96US-0630048.

PR 06-MAR-1997; 97US-0038614.

PA (BARD-) BARD DIAGNOSTIC SCI INC.

PI Enfield DL, Hass GM, Kinders RJ;

DR WPI; 1997-512742/47.

DR N-PSDB; V02791.

PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
 PT by modulating or detecting tumour associated human complement factor  
 PT H related antigen, or nucleic acid encoding it

XX  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein is found in clone PRB9FH410 and represents a  
 CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.  
 CC  
 SQ Sequence 216 AA.

Query Match 65.5%; Score 1229; DB 18; Length 216;  
 Best Local Similarity 99.5%; Pred. No. 9.3e-86;  
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 95 TCNEGQQLLGETINRECDTGDWNTDIPICEVVKCLPVTAPENGKIVSSAMEPDREYHFGQ 154  
 DB 1 tcnegqyllgetnrecdtdgwtndipicevkcldpvtapengkivsamedpreyhfg 60  
 QY 155 AVAFVCSNGYKTEGDEMHCSDDGFWSKPKVCVEISCKSPDVINGSPISOKIITKENER 214  
 DB 61 avrfvcsngyklegdeemhcsddgfwskpkvcveisckspdvingspsiskilykener 120  
 QY 215 FOYKCNMGYEYSEKDAVCTESGMRPLPSCCEKSCDNPYIPNGDYSPLRIKHTGDEITY 274  
 DB 121 foykcnmgysersgavctesgmrplpsceekscdnpyipngdysplrikhtgdeity 180  
 QY 275 QCRNGYTPATRGNTAKCTSTGIPAPRCTLKPCDYP 310  
 DB 181 qcrngytpatrgntakctstgwaprccltkpcdyp 216

## RESULT 3

Y55752  
 ID Y55752 standard; Protein: 496 AA.

AC Y55752;

DT 22-FEB-2000 (first entry)

DE Human CRI protein LHR-A SCR fragment.

KW C3B/C4b receptor; CRI protein; cell-surface protein; erythrocyte; human;  
 KW complement regulatory activity; complement pathway enzyme; tissue damage;  
 KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;  
 KW heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;  
 KW short consensus repeat.

OS Homo sapiens.

PN US5981481-A.

PD 09-NOV-1999.

PE 06-JUN-1995; 95US-0470652.

PR 03-APR-1989; 89US-0332865.

PR 06-DEC-1974; 74US-0350238.

PR 24-FEB-1993; 93US-0026134.

PR 01-APR-1988; 88US-0176532.

PA (UYUO ) UNIV JOHNS HOPKINS.

PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.

PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.

PI Conclino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;  
 PI Marsh HC, Carson GR;  
 DR WPI; 1999-633357/54.

Query Match	Best Local Similarity	Score	DB	Length
Matches 102; Conservative	18.8%;	353;	DB 20;	496;
	29.8%;	Pred. No. 3.7e-19;		
	54;	Mismatches 126;	Indels 60;	Gaps 24
3 CNE---	LPRRMTIELTGSWSDQTYDEGNQATYKCRPGRSIGNVIMCRKEG-WVALNP	58		
48	capewlptarpcnlt---	defetpigtlylneyccpysgtrpfslcclnkswtgacd	103	
59	LKKCQKRPCHGHDPEFGFTLLGNVFEYGVKAVYTCMEGYOLLGEINTEC---	DTD 114		
104	--rrrrkscrrppdpnqgmhwkfg--	iqfsgqikysctckqylligs--saatclisgdtv	158	
115	GWTNDIPICEVYKCC-LPVTAPENGKATVSSAMERDRXHHGQARFQCNSS-----	YKIE 167		
159	lvndneplcdirlpcgllprclt-lngdfisthre--	nlhysvvtlyrcmpsgsgtrkvwfelv	214	
168	GDEEMHC--SD--GHSKEKPKC-VEISCKSPDVINGSPIS-QKIIYKENERFQYKCM	221		
215	gepslyctsnddvgvfwspapqcllnpctprvengllvsnrlsfslsnevefrcqp	274		
222	GVEYSERGAVCTE-SGWRP-LPSCEEKSCDNPYIPN-----	GDYSPLRIKHRT 268		
275	gfymkqprtrvkcgqalnkwpelpsc-sryccqpP--	pdlvhaertqrldkdnfsp-----	324	
269	GDEITTYQCRNGFYPATRGNTAKCTSTG-WIP-APRCTLPKPCD	308		
325	ggevtyscpepg-ydlrgaasmrctpgdgdwspaatpcevkscd	365		
RESULT 4				
ID	RL3490			
AC	RL3490 standard; Protein; 581 AA.			
XX				
XX	30-OCT-1991 (first entry)			
XX				
XX	Human C4 binding protein.			
XX				
KM	C4bp; monomer; complement protein; pJOD.C4bp.3; SCR;			
XX	short consensus repeat.			
OS	Homo sapiens.			
XX				
FH	Key	location/Qualifiers		
FT	Peptide	1..32		
FT	Protein	/label= signal_peptide		
FT		33..581		
FT		/label= C4bp		

FT	Region	33...93	/label= SCR8
FT	Region	94...155	/label= SCR7
FT	Region	156...219	/label= SCR6
FT	Region	220...279	/label= SCR5
FT	Region	280...345	/label= SCR4
FT	Region	346...406	/label= SCR3
FT	Region	407...464	/label= SCR2
FT	Region	465...523	/label= SCR1
FT	Domain	524...581	/label= C4bp_core
FT		/note= "responsible for multimer assembly"	
FT	Disulfide-bond	34...80	/note= "Intradomain"
FT	Disulfide-bond	65...92	/note= "Intradomain"
XX	W09111461-A.		
XX	08-AUG-1991.		
XX	28-JAN-1991;	91WO-US00567.	
XX	26-JAN-1990;	90US-0470888.	
XX	(BIOG-) BIOGEN INC.		
XX	Pasek MP, Winkler G, Liu TR;		
XX	WPI: 1991-252613/34.		
XX	N-PSDB: Q13242.		
PT	New C4 binding protein fusion proteins and DNA encoding them -		
PT	comprise assemblies of C4bp monomers linked to functional moiety,		
PT	e.g. A2T, useful as delivery vehicles in diagnosis and therapy		
PS	Example 1; Fig 1; 105pp; English.		
CC	This sequence was deduced from human hepatocyte (Hep G2) cDNA		
CC	obtained following PCR amplification. The protein is a monomer		
CC	containing 8 SCRs. Each SCR forms a looped domain due to the		
CC	presence of two intradomain disulphide bonds (only the disulphide		
CC	bonds of SCR8 are labelled in the Features Table). Within each SCR,		
CC	the first cysteine residue bonds with the third and the second		
CC	cysteine residue bonds with the fourth. This secondary structure is		
CC	responsible for the conformational flexibility of the C4bp monomer.		
CC	The invention covers fusion proteins in which the monomer sequence,		
CC	or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCRs(s)		
CC	is used to the C-terminal of a protein such as a viral receptor,		
CC	cell ligand, a bacterial, viral or parasitic immunogen, enzyme,		
CC	cytokine, toxin, etc. See also Q13243-51.		
XX	Sequence	581 AA:	
Query Match	18.8%;	Score 353;	DB 12; Length 581;
Best Local Similarity	29.3%;	Pred. No. 4,4e-19;	
Matches 100;	Conservative 50;	Mismatches 141;	Indels 50; Gaps 21.
QY	20	SDQTYEGGTAIKKCPGY-RSLGNVIMCRK-GEWVALNPLKCOKRCGHPGDPEFT	77
	::::		
DB	51	tetrlkgtlkytclpypyrnsqticnsdgew-yntf--clykrcrlpcewv-----	102
QY	78	FTLTGGNV-----FEYGVKAVYTCNCGYOLLGEINRECDTD---GWTNDIPICEVVC	128
		:::	
DB	103	--lringvvelktdltsfgsqfescseggfllgltar-cevqdraygwahplpgecivkc	159

```

OY 129 LPVTAPENGKIYSSAMEPPREYHFGQAVRVCNSGKIEDEDMHCSDD---GFMSKEK 184
DB 160 kppdltmrnrhsg---enfyaygitsvyscdprfslilghaslsctvenetlgvtrp 215
OY 185 PKCEVISCSPDINGSPISQ--KIYKENERFOYKCNMGCEYSEKGDV--C-TESGWRP 240
DB 216 pccckltcrkpdvshgmsvsgfgrlynykdlvfkckqkfvl--rsgsvlhcddaskmp 273
OY 241 L-PSCEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITYOCRNCPYPAT--RGNTAKC- 291
DB 274 sspacecnscllppdlphaswelyprtkedyyvgtvlrtrchpkykrltdeptviciq 333
OY 292 TSTGWIPARCTLKPCDYDPIKHGGLYHNMRRP-----YF 327
DB 334 klrtvtrpygcealccpepklnngeltqhrksrpanhcvyf 374

RESULT 5
R28570
ID R28570 standard; peptide: 543 AA.
XX
AC R28570;
XX
DT 19-MAR-1993 (first entry)
XX
DE CRI-4 (266-274 KLKTQTNASD) analogue.
XX
KW short consensus repeat; regulator of complement activation;
KM C3b binding; C4b binding; human complement type 1 receptor.
XX
OS Homo sapiens.
XX
FH Key location/Qualifiers
FT Region 1..60
FT Region /label= SCR-1
FT Region 61..122
FT Region /label= SCR-2
FT Region 451..510
FT Region /label= SCR-8
FT Region 511..543
FT Region /label= SCR-9
FT Region /note= "TRUNCATED"
FT MISC-difference 266..274
FT /note= "ERTQORDKN substituted with KLKTQTNASD
from SCR-8-9 to increase C3b binding"
XX
PD EP512733-A.
XX
PD 11-NOV-1992.
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
XX Atkinson JP, Hourcade D, Krych M;
XX
XX WPI; 1992-375009/46.
XX
XX
XX Complement activity regulator protein analogues - useful for
XX treating auto-immune diseases, to suppress transplant rejection,
XX for diagnosis etc.
XX
XX Example 8; Page 18 and R11810; 23pp; English.
XX
XX The CDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX 168:1255-1270. It encodes the first 8 and a half amino terminal
XX SCRs of CRI. The invention concerns analogues of "regulator of
XX complement activation" proteins or truncated, hybrid or recombinant
XX forms of them. CRI-4 is a preferred truncated form and a number of
XX specified substitution variants of it are disclosed in which certain

```

```

CC positions in SCR-5-6 are substituted by amino acids from
CC the corresponding positions in SCRs which are involved in C3b- and
CC C4b-binding. The substitution variant given here has increased C3b-
CC binding. The specification does not contain the CRI-4 sequence;
CC the sequence given here was constructed from the full-length CRI
CC amino acid sequence having GENESEQ accession number R11810 and
CC descriptions in the disclosure.
XX
SQ Sequence 543 AA:
XX
Query Match 18.8%; Score 352.5; DB 13; Length 543;
Best local Similarity 29.6%; Pred. No. 4,5e-19;
Matches 101; Conservative 53; Mismatches 128; Indels 59; Gaps 23;
OY 3 CNE---LPPRAMEILTGMSDQTYPEGNOATYKCRPGYRSLGNVIMWCRKGWALNPL 59
DB 2 cnapewlpfartnlt----delefplgtylnyecrpyggrpfslclksvwtgkd- 56
OY 60 RKCQKRRCGHPDTPGTEFLTGAVFEYKAVYVCNMGYQLAGEINREC---DTDG 115
DB 57 -rcrtksctnrpdpngmwhvlg--lqfsgqikysctkyrligs-satclsgdvtl 112
OY 116 WTNDIPICEVVC-LPVTAPENGKIYSSAMEPPREYHFGQAVRVCNSG-----YKIEG 168
DB 113 wdnefpicdrlpogilppl-tngdfistnre--nfhygsvvtlyrcnpgsggrkvtelvg 168
OY 169 DEEMHC--SDD--GFMSKRPKC-VEISCKSPDINGSPIIS-QKIYKENERFOYKCNNG 222
DB 169 epslyctsnddqvgjlsgapqcelipnktcpvneglivsdnrlfslnevefrcqpv 228
OY 223 YEYSEKDAVCTE-SGWRP-LPSCCEKSCDNPIIP-----NGDISPLRIKHRTG 269
DB 229 fymkgprtyvcqalnkwepelpsc-srvccqp--pdvlnakikqtusdpsf-----g 278
OY 270 DEITYOCRNCFYPATRGNTAKTSTG-WIP-APRCTIKPCD 308
DB 279 qevfyscepg-ydlrtaasmrctpgdwsapaftcevxscd 318

RESULT 6
W73147
ID W73147 standard; protein: 778 AA.
XX
AC W73147;
XX
DT 29-JAN-1999 (first entry)
XX
DE Amino acid sequence of the soluble complement receptor 1 (SCR1).
XX
XX
XX Human: soluble complement receptor 1; SCR1; T-cell; B-cell;
XX mediated immune response; inhibition; tissue rejection; gene therapy;
XX dystrophin; inflammatory response; interferon-gamma secretory response;
XX autoimmune response; neurological response; Alzheimer's disease;
XX Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;
XX rheumatoid arthritis; myasthenia gravis; epidermis bullousa;
XX Hashimoto's disease.
XX
XX Homo sapiens.
XX
XX W09845430-A1.
XX
XX 15-OCN-1998.
XX
XX 06-APR-1998; 98WO-GB01012.
XX
XX 05-APR-1997; 97GB-0006950.
XX
XX (ANNE/) ANNENKOV A.
XX (CHER/) CHERNAJOVSKY Y.
XX
XX Annenkov A, Chernajovsky Y;
XX

```



DR WPI: 1998-568350/48.  
 XX Fragment of soluble human complement receptor 1 - useful for  
 PT treating T-cell or B-cell mediated immune responses e.g.  
 PT inflammatory responses such as rheumatoid arthritis  
 PS Disclosure; Fig 1: 54pp; English.  
 XX  
 XX This is an amino acid sequence of the human soluble complement  
 CC receptor 1 (SCR1), useful in the treatment of T-cell or B-cell  
 CC mediated immune responses. It is used to inhibit a T-cell or  
 CC B-cell mediated immune response to prevent immune response-mediated  
 CC tissue rejection and destruction or clearance or inactivation of an  
 CC expressed protein especially from cells that have been treated by gene  
 CC therapy to express the protein, e.g. dystrophin. The protein can also  
 CC be used to inhibit a T-cell or B-cell-mediated inflammatory response,  
 CC an interferon-gamma secretory response, autoimmune response or  
 CC neurological response, e.g. Alzheimer's or Parkinson's disease or  
 CC multiple sclerosis. Also the protein can be used to treat systemic  
 CC lupus erythematosus, rheumatoid arthritis, myasthenia gravis,  
 CC epidermis bullosa or Hashimoto's disease.  
 CC  
 XX Sequence 778 AA:  
 SO  
 Query Match 18.7%; Score 350.5; DB 19; Length 778;  
 Best Local Similarity 29.3%; Pred. No. 9.7e-19;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRGRYSGLGNVIMVCKRKGWALNPL 59  
 DB 48 cnapewlfpairpnl-----defefpigtlylneyecrpysgrpfslclksnvtgacd- 102  
 QY 60 RKCKRPGCHPGDTPFGTFTLTGWNVEYGVKAVYTCNCEYOLLGEINREK---DTDG 115  
 DB 103 -rcrkscrrppdvngmwhvlykg--lqfsgqkysctkyrligs-saatcllsgdtvla 158  
 QY 116 WTNDIPICEVVK-LPVTAPENKIVSAMPEPDREYHFGQAVFVNCNSG-----YKIEG 168  
 DB 159 wdnetpictdipcgllppl-tngdfistnre--nfhygsvvtlyrcnpgsgqrkfvclvg 214  
 QY 169 DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222  
 DB 215 epslyctsndqygiwspapqclipmkcprpvenngilvsdnstlsinevelfrcpgy 274  
 QY 223 YEYSERGDVCTE-SGMRP-LPSCSEKSCDNPIYIPN-----GDYSPRIKIHRTG 269  
 DB 275 fvmkgprtrvkqalnwepelpsc-srvccpp--pdvllaertgrdkdnfisp-----g 324  
 QY 270 DEITYOCRNNGFYPATRGNTAKCTSTG-WIP-APRCTLAKPCD 308  
 DB 325 qevfyscepg-ydlrngaasmrcrpgdwsapaclcevkscd 364  
 Db  
 RESULT 7  
 W43899 ID W45899 standard; peptide; 1930 AA.  
 XX W45899;  
 XX  
 XX 30-JUN-1998 (first entry)  
 XX  
 XX Human complement receptor 1 (residues 1-1929).  
 XX  
 XX Membrane binding element; thrombotic disease; soluble protein;  
 KW complement-related disease; integral membrane protein; inflammation;  
 KW short consensus repeat; SCR 1-3; CRI; complement receptor type 1.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 1930  
 FT Cross-Links  
 FT /note= "Disulphide linked to Cys in peptide given

FT in W45889"  
 XX W09802454-A2.  
 XX 22-JAN-1998.  
 PD  
 XX 08-JUL-1997; 97WO-EP03715.  
 PF  
 XX 15-JUL-1996; 96GB-0014871.  
 PR  
 XX (ADPR-) ADPROTECH PLC.  
 PA  
 PI Dodd I, Mossakowska DEL, Smith RAG;  
 DR WPI: 1998-110524/10.  
 XX  
 XX Derivatives of soluble poly(peptide)s bonded to low affinity  
 PT membrane binding groups - useful for treating complement-related and  
 PT thrombotic diseases, providing improved localisation at cellular  
 PT membranes  
 PS  
 XX Claim 22; Pages 60-61; 75pp; English.  
 CC  
 CC This sequence represents human complement receptor 1 (CRI, CD 35)  
 CC N-terminal fragment. The invention relates to a soluble derivative (A)  
 CC of a soluble polypeptide (I), which comprises at least 2 heterologous  
 CC membrane-binding elements (MBE) of low membrane affinity covalently  
 CC associated with (I). MBE interact, independently and with thermodynamic  
 CC additivity, with components of cellular or artificial membranes exposed  
 CC to extracellular fluids. (A) are used to treat disorders treatable with  
 CC (I) itself, specifically inflammation or any other complement-related  
 CC disorder (e.g. neurological disease, graft rejection, myocardial  
 CC infection, sepsis, rheumatoid arthritis and many others; including  
 CC application to indwelling devices) and thrombolytic disease, but also to  
 CC treat allergy, induce weight loss, to treat ischaemia or asthma and as  
 CC immuno-modulators for treating multiple sclerosis. (A) are administered  
 CC orally, topically, by injection or inhalation at 0.01-10 (preferably  
 CC 0.1-10) mg/kg/day.  
 CC  
 XX Sequence 1930 AA:  
 SQ  
 Query Match 18.7%; Score 350.5; DB 19; Length 1930;  
 Best Local Similarity 29.3%; Pred. No. 2.9e-18;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;  
 QY 3 CNE---LPPRRNTEILTGSMSDQTYPEGTQAIYKCRGRYSGLGNVIMVCKRKGWALNPL 59  
 DB 2 cnapewlfpairpnl-----defefpigtlylneyecrpysgrpfslclksnvtgacd- 56  
 QY 60 RKCKRPGCHPGDTPFGTFTLTGWNVEYGVKAVYTCNCEYOLLGEINREK---DTDG 115  
 DB 57 -rcrkscrrppdvngmwhvlykg--lqfsgqkysctkyrligs-saatcllsgdtvla 112  
 QY 116 WTNDIPICEVVK-LPVTAPENKIVSAMPEPDREYHFGQAVFVNCNSG-----YKIEG 168  
 DB 113 wdnetpictdipcgllppl-tngdfistnre--nfhygsvvtlyrcnpgsgqrkfvclvg 168  
 QY 169 DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDVINGSPIS-OKIYKENERFOYKCNMG 222  
 DB 169 epslyctsndqygiwspapqclipmkcprpvenngilvsdnstlsinevelfrcpgy 228  
 QY 223 YEYSERGDVCTE-SGMRP-LPSCSEKSCDNPIYIPN-----GDYSPRIKIHRTG 269  
 DB 229 fvmkgprtrvkqalnwepelpsc-srvccpp--pdvllaertgrdkdnfisp-----g 278  
 QY 270 DEITYOCRNNGFYPATRGNTAKCTSTG-WIP-APRCTLAKPCD 308  
 DB 279 qevfyscepg-ydlrngaasmrcrpgdwsapaclcevkscd 318  
 Db  
 RESULT 8  
 Y55751

[illegible]

Db	154	wdhnetpictdrlpoglplti.tngdifistnre---nfiygsavlyrcmpsggikvfeivg	209
Qy	169	DEEMHC--SDD--GEWSKEKPKC-VEISCKSPDIVINGSPIS-QRIYKENERFOYKCMWG	222
Db	210	epsistcndddqygiwsgpapqclipnkcicpnuvenqllvsdnrlsfalsnevvefrcpg	269
Qy	223	YKESERGDVACTE-SGWRP-LPSCEEKSCONPIYIPN-----GDYSEPLRIKHRTG	269
Db	270	fmvkgprvkvkcgalknkepelpsc-strvcgpp--pdvlnaertlrkdxndfisp-----g	319
Db	320	gevfyscepg-ydlrgaasmrctpcpgdwpspaaprcvkscd	359
RESULT	9		
ID	P92219	standard; protein; 2317 AA.	
AC	P92219;		
XX	22-FEB-1990	(first entry)	
DT	CRI protein.		
XX	Complement; cofactor.		
OS	Homo sapiens (human).		
XX	Key	Location/Qualifiers	
FT	Peptide	10..50	
XX		/label= signal_peptide	
PN	W08909220-A.		
PD	05-OCT-1989.		
XX	31-MAR-1989;	89WO-0501358.	
PR	01-APR-1988;	88US-0176532.	
XX	(TCHEL ) T CELL SCIENCES INC.		
PA	(UYIO ) THE JOHNS HOPKINS UNIVERSITY.		
XX	(BRIG ) THE BRIGHAM AND WOMEN'S HOSPITAL.		
PI	Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;		
DR	WPI; 1989-309498/42.		
DR	N-PSDB; N91477.		
XX	New nucleic acid sequences encoding new CRI protein - and its fragment,		
PT	for diagnosis and control of complement-related immune defects,		
PT	inflammation, myocardial infarct, etc		
XX	Claim 1; Fig 1; 191pp; English.		
XX	This is full-length CRI protein, and shortened forms are new, lacking		
CC	the transmembrane region. The proteins and fragments bind C3b and/or		
CC	C4d, have cofactor activity and inhibit C3 and C5 convertase activity.		
CC	In the sequence, x=untranslated region. This has 7 short consensus		
CC	repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in		
CC	patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C		
CC	for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.		
CC	They are useful in diagnosing and treating immune disorders, and prevent		
XX	perfusion injury.		
XX	Sequence	2317 AA;	
XX	Query Match	18.7%;	Score 350.5; DB 10; Length 2317;
XX	Best Local Similarity	29.3%;	Pred. No. 3,6e-18;
XX	Matches 100; Conservative	55; Mismatches 127;	Indels 59; Gaps 23



```

FT      /note= "Thr substituted by Ser (SCR-8)"
FT      Misc-difference 53
FT      /note= "Gly substituted by Ser (SCR-8)"
FT      Misc-difference 54
FT      /note= "Ala substituted by Pro (SCR-8)"
XX      EP512733-A.
XX      11-NOV-1992.
XX      28-APR-1992; 92EP-0303826.
XX      03-MAY-1991; 91US-0695514.
XX      (UNIM ) UNIV WASHINGTON.
XX      Atkinson JP, Hourcade D, Krych M;
XX      WPI; 1992-375009/46.
XX      Complement activity regulator protein analogues - useful for
XX      treating auto-immune diseases, to suppress transplant rejection,
XX      for diagnosis etc.
XX      Claim 11; Fig 2 and R11810; 23pp; English.
XX      The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX      168:1255-1270. It encodes the first 8 and a half amino terminal
XX      SCRs of CRI. The invention concerns analogues of "regulator of
XX      complement activation" proteins or truncated, hybrid or recombinant
XX      forms of them. CRI-4 is a preferred truncated form and a number of
XX      specified substitution variants of it are claimed in which certain
XX      positions in SCR-1 which have been identified as important for the
XX      degree of C3b- and C4b-binding are substituted by amino acids from
XX      the corresponding positions in SCR-8. The specification does not
XX      contain the CRI-4 sequence; the sequence given here was constructed
XX      from the full-length CRI amino acid sequence having GENESEQ
XX      accession number R11810 and descriptions in the disclosure.
XX      Sequence 543 AA;

Query Match 18.4%; Score 345.5; DB 13; Length 543;
Best Local Similarity 29.3%; Pred. No. 1.5e-18;
Matches 100; Conservative 56; Mismatches 126; Indels 59; Gaps 23;

QY      3 CNE---LPPRRNTEILGSMDSQTYREGTOAITKCRGYSGLGNVWCKRGEWALNPL 59
QY      ||||| : : : : : ||||| : : : : :
QY      2 cnapewlpfarpnl-----defefpigtlynyecrpgysgrfslclknsw--sspk 55
QY      60 RKCQRPCGHPGDPPTFTLTGNGWFEYGVKAVYTCNEGVLGELINRYRC---DTDG 115
QY      ||||| : : : : : ||||| : : : : :
QY      56 drcrkscrrppdvngmvnviqg--iqfsgqlyskctkyrlligs--ssatclisgdtvi 112
QY      116 WTNDIPICEVKK-LPYTADENKIVSAMPEDEYHFGQAVRFVCSNG-----YKIEG 168
QY      ||||| : : : : : ||||| : : : : :
QY      113 whetpictdrlpcglppli--tngfistlne---nfhygsvlytrcnpqsggrkvfelvg 168
QY      169 DEEMHC--SDD--GFWGKERPKC-VEIISCKSPVYINSPLS--QKTIYKENDRFRQKCNMG 222
QY      ||||| : : : : : ||||| : : : : :
QY      169 epsiylcsnddgyqivswpapqcllpnkcptpvevgilvsdnslfsleuvefrcqpv 228
QY      223 YEYSEKGDAACTE-SGRP-LPSCEEKSCDNPIYPN-----GDSPLRIKHRTG 269
QY      ||||| : : : : : ||||| : : : : :
QY      229 fymkgprtrvvcgqlnkwpelpsc--stvcqpp--pdvlhaertqrkdndisp-----g 278
QY      270 DETTYOCRRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPDC 308
QY      ||||| : : : : : ||||| : : : : :
QY      279 qevfyscepg-ydlrgaasmrctcpqgdwspapctcevkscd 318

```

```

ID      R28560 standard; peptide; 543 AA.
XX      AC R28560;
XX      19-MAR-1993 (first entry)
XX      CRI-4 (114S) analogue.
XX      short consensus repeat; regulator of complement activation;
XX      C3b binding; C4b binding; human complement type 1 receptor.
XX      Homo sapiens.
XX      Key Location/Qualifiers
XX      Region 1..60
XX      Region /label= SCR-1
XX      Region 61..122
XX      Region /label= SCR-2
XX      Region 451..510
XX      Region /label= SCR-8
XX      Region 511..543
XX      /label= SCR-9
XX      /note= "TRUNCATED"
XX      Misc-difference 114
XX      /note= "Asp substituted by Ser (SCR-9)"
XX      EP512733-A.
XX      11-NOV-1992.
XX      28-APR-1992; 92EP-0303826.
XX      03-MAY-1991; 91US-0695514.
XX      (UNIM ) UNIV WASHINGTON.
XX      Atkinson JP, Hourcade D, Krych M;
XX      WPI; 1992-375009/46.
XX      Complement activity regulator protein analogues - useful for
XX      treating auto-immune diseases, to suppress transplant rejection,
XX      for diagnosis etc.
XX      Claim 11; Fig 2 and R11810; 23pp; English.
XX      The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX      168:1255-1270. It encodes the first 8 and a half amino terminal
XX      SCRs of CRI. The invention concerns analogues of "regulator of
XX      complement activation" proteins or truncated, hybrid or recombinant
XX      forms of them. CRI-4 is a preferred truncated form and a number of
XX      specified substitution variants of it are claimed in which certain
XX      positions in SCR-2 which have been identified as important for the
XX      degree of C3b- and C4b-binding are substituted by amino acids from
XX      the corresponding positions in SCR-9. The specification does not
XX      contain the CRI-4 sequence; the sequence given here was constructed
XX      from the full-length CRI amino acid sequence having GENESEQ
XX      accession number R11810 and descriptions in the disclosure.
XX      Sequence 543 AA;

Query Match 18.3%; Score 343.5; DB 13; Length 543;
Best Local Similarity 29.0%; Pred. No. 2.1e-18;
Matches 99; Conservative 56; Mismatches 127; Indels 59; Gaps 23;

QY      3 CNE---LPPRRNTEILGSMDSQTYREGTOAITKCRGYSGLGNVWCKRGEWALNPL 59
QY      ||||| : : : : : ||||| : : : : :
QY      2 cnapewlpfarpnl-----defefpigtlynyecrpgysgrfslclknswtqakd- 56
QY      60 RKCQRPCGHPGDPPTFTLTGNGWFEYGVKAVYTCNEGVLGELINRYRC---DTDG 115
QY      ||||| : : : : : ||||| : : : : :
QY      57 -rcrkscrrppdvngmvnviqg--iqfsgqlyskctkyrlligs--ssatclisgdtvi 112

```



```

XX 03-MAY-1991; 91US-0695514.
XX (UNITW ) UNIV WASHINGTON.
XX Atkinson JP, Hourcade D, Krych M;
XX WPI: 1992-375009/46.
XX
XX Complement activity regulator protein analogues - useful for
XX treating auto-immune diseases, to suppress transplant rejection,
XX for diagnosis etc.
XX
XX Claim 11; Fig 2 and R11810; 23pp; English.
XX
XX The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX 168:1255-1270. It encodes the first 8 and a half amino terminal
XX SCRs of CRI. The invention concerns analogues of "regulator of
XX complement activation" proteins or truncated, hybrid or recombinant
XX forms of them. CRI-4 is a preferred truncated form and a number of
XX specified substitution variants of it are claimed in which certain
XX positions in SCR-2 which have been identified as important for the
XX degree of C3b- and C4b-binding are substituted by amino acids from
XX the corresponding positions in SCR-9. The specification does not
XX contain the CRI-4 sequence; the sequence given here was constructed
XX from the full-length CRI amino acid sequence having GENESEQ
XX accession number R11810 and descriptions in the disclosure.
XX
XX Sequence 543 AA:
XX
XX Query Match 18.2%; Score 341.5; DB 13; Length 543;
XX Best Local Similarity 29.0%; Pred. No. 3e-18;
XX Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;
XX
XX QY 3 CNE---LPPRNTTEILGWSMSDQTYPECTOAIYKCRGYSRLGNVIMCRKGEWALNPL 59
XX || || | : : : : : || || || || || : : |
XX 2 cnapewlpfarptnlt-----defetpigtlylneecrpygsrpfslclksnvtgskd- 56
XX
XX QY 60 RKCKRPGHGDPTFGFTLTGWNVEYKAVYTCNEGVLGEINVRCC---PTDG 115
XX || || | : : : : : || || || || || : : |
XX 57 -rrctkscrnppdpvngmwnhikg--lqfsgqlkyctkgyrlilgs--ssatclisgdtvi 112
XX
XX QY 116 WTNDIPICEVVKC-LPYTAPENKIVSSAMPEDEYHFGQAVRFVNCNSG-----YKIEG 168
XX || || | : : : : : || || || || || : : |
XX 113 wdnetpictqrlpcqlppll-tngdflsture---nfhygsvvtlyrcnpgsggrkvfelvg 168
XX
XX QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDYINSGPIS--QKIIYKENERFOYKCNMG 222
XX || || | : : : : : || || || || || : : |
XX 169 epslyctsnddqygiwsgpapqcllpnkctcpnvengllvsdnrlslsinevvefrcpv 228
XX
XX QY 223 YEYSEKDAVCTE--SGWRP--LPSCSEKSCDNPYIPN-----GDVSPRLRIKHRTG 269
XX || || | : : : : : || || || || || : : |
XX 229 fymkgprlvkqgalnkweelpsc--stvcqpp--pdlhaerctgrdkdnisp-----g 278
XX
XX QY 270 DEITYOQRNGFYFATRGNTAKCTSTG-WIP-APRCTLTKPCD 308
XX || || | : : : : : || || || || || : : |
XX 279 gevfyscepg-ydlrgaasmctcpqgdwspaprcvksd 318
XX
XX RESULT 15
XX R28565
XX ID R28565 standard; peptide; 543 AA.
XX
XX AC R28565;
XX
XX DT 19-MAR-1993 (first entry)
XX
XX DE CRI-4 (1210) analogue.
XX
XX KW short consensus repeat; regulator of complement activation;
XX C3b binding; C4b binding; human complement type 1 receptor.
XX
XX

```

```

OS Homo sapiens.
XX Key Location/Qualifiers
XX Region 1..60
XX FT /label= SCR-1
XX Region 61..122
XX FT /label= SCR-2
XX Region 451..510
XX FT /label= SCR-8
XX Region 511..543
XX FT /label= SCR-9
XX FT /note= "TRUNCATED"
XX FT Misc-difference 121
XX FT /note= "Asp substituted by Gln (SCR-9)."
XX
XX EP512733-A.
XX
XX 11-NOV-1992.
XX
XX 28-APR-1992; 92EP-0303826.
XX
XX 03-MAY-1991; 91US-0695514.
XX
XX (UNITW ) UNIV WASHINGTON.
XX
XX Atkinson JP, Hourcade D, Krych M;
XX WPI: 1992-375009/46.
XX
XX Complement activity regulator protein analogues - useful for
XX treating auto-immune diseases, to suppress transplant rejection,
XX for diagnosis etc.
XX
XX Claim 11; Fig 2 and R11810; 23pp; English.
XX
XX The cDNA clone designated CRI-4 was described in J.Exp.Med.(1988)
XX 168:1255-1270. It encodes the first 8 and a half amino terminal
XX SCRs of CRI. The invention concerns analogues of "regulator of
XX complement activation" proteins or truncated, hybrid or recombinant
XX forms of them. CRI-4 is a preferred truncated form and a number of
XX specified substitution variants of it are claimed in which certain
XX positions in SCR-2 which have been identified as important for the
XX degree of C3b- and C4b-binding are substituted by amino acids from
XX the corresponding positions in SCR-9. The specification does not
XX contain the CRI-4 sequence; the sequence given here was constructed
XX from the full-length CRI amino acid sequence having GENESEQ
XX accession number R11810 and descriptions in the disclosure.
XX
XX Sequence 543 AA:
XX
XX Query Match 18.2%; Score 341.5; DB 13; Length 543;
XX Best Local Similarity 29.0%; Pred. No. 3e-18;
XX Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;
XX
XX QY 3 CNE---LPPRNTTEILGWSMSDQTYPECTOAIYKCRGYSRLGNVIMCRKGEWALNPL 59
XX || || | : : : : : || || || || || : : |
XX 2 cnapewlpfarptnlt-----defetpigtlylneecrpygsrpfslclksnvtgskd- 56
XX
XX QY 60 RKCKRPGHGDPTFGFTLTGWNVEYKAVYTCNEGVLGEINVRCC---PTDG 115
XX || || | : : : : : || || || || || : : |
XX 57 -rrctkscrnppdpvngmwnhikg--lqfsgqlkyctkgyrlilgs--ssatclisgdtvi 112
XX
XX QY 116 WTNDIPICEVVKC-LPYTAPENKIVSSAMPEDEYHFGQAVRFVNCNSG-----YKIEG 168
XX || || | : : : : : || || || || || : : |
XX 113 wdnetpictqrlpcqlppll-tngdflsture---nfhygsvvtlyrcnpgsggrkvfelvg 168
XX
XX QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDYINSGPIS--QKIIYKENERFOYKCNMG 222
XX || || | : : : : : || || || || || : : |
XX 169 epslyctsnddqygiwsgpapqcllpnkctcpnvengllvsdnrlslsinevvefrcpv 228
XX
XX QY 223 YEYSEKDAVCTE--SGWRP--LPSCSEKSCDNPYIPN-----GDVSPRLRIKHRTG 269
XX || || | : : : : : || || || || || : : |
XX

```

Db 229 fvmkgprvkcgalnkwepe1psc-srvccpp--pdv1haertgrdkdnfsp-----g 278  
Oy 270 DEITYOCRNNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308  
| : | | | | : | | | | : | |  
| : | | | | : | | | | : | |  
Db 279 gevfysecpq-ydlrgaasmrctpgdwsapaapcevkscd 318

Search completed: November 21, 2000, 16:43:48  
Job time: 287 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:19 ; Search time 27.94 Seconds  
(without alignments)  
303.014 Million cell updates/sec

Title: US-09-316-163-10  
Perfect score: 1497  
Sequence: 1 EDCNELPPRRNTTEILTGSWS.....EKSCDNPYIPNGDYSPLRIK 265

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1497	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1024	68.4	1234	1 CFAH_MOUSE	P06909 mus musculu
3	303	20.2	597	1 C4BP_HUMAN	P04003 homo sapien
4	301.5	20.1	263	1 VCP_VACCV	P10998 vaccinia vi
5	298.5	19.9	2039	1 CRI_HUMAN	P17927 homo sapien
6	276.5	18.5	558	1 C4BP_MOUSE	P08607 mus musculu
7	275	18.4	459	1 C4BP_MOUSE	P15529 homo sapien
8	274	18.3	377	1 MCP_HUMAN	P15529 homo sapien
9	272	18.2	610	1 LEM2_HUMAN	P08110 sus scrofa
10	266	17.8	484	1 LEM2_PIG	Q28065 bos taurus
11	265.5	17.7	610	1 C4BP_BOVIN	Q01016 herpesvirus
12	264.5	17.7	360	1 C4BP_HSVSA	P08174 homo sapien
13	262	17.5	381	1 DAF_HUMAN	P00102 mus musculu
14	260.5	17.4	768	1 LEM3_MOUSE	P00690 mus musculu
15	260	17.4	612	1 LEM2_MOUSE	P19070 mus musculu
16	257.5	17.2	1025	1 CR2_MOUSE	P08106 rattus norv
17	254	17.0	768	1 LEM3_RAT	P20023 homo sapien
18	254	17.0	1033	1 CR2_HUMAN	Q01333 mus musculu
19	253	16.9	345	1 APOH_MOUSE	P49457 pongo pygma
20	252	16.8	340	1 DAF_PONPY	P05160 homo sapien
21	252	16.8	661	1 F13B_HUMAN	P16109 homo sapien
22	250.5	16.7	830	1 LEM3_HUMAN	P08109 ovis aries
23	249.5	16.7	769	1 LEM3_SHEEP	Q06401 cavia porce
24	246.5	16.5	507	1 DAF_CAVPO	P06105 rattus norv
25	241	16.1	549	1 LEM2_RAT	P33730 canis fami
26	241	16.1	611	1 LEM2_CANFA	P17690 bos taurus
27	240	16.0	345	1 APOH_BOVIN	P27113 oryctolagus
28	239.5	16.0	551	1 LEM2_RABIT	P00968 mus musculu
29	234	15.6	668	1 F13B_MOUSE	Q09101 drosophila
30	233.5	15.6	958	1 HIG_DROME	Q01475 mus musculu
31	231.5	15.5	390	1 DAF1_MOUSE	P07749 homo sapien
32	231	15.4	345	1 APOH_HUMAN	Q06146 mus musculu
33	222.5	14.9	407	1 DAF2_MOUSE	

34	222.5	14.9	485	1 LEM2_BOVIN	P08107 bos taurus
35	222	14.8	345	1 APOH_CANFA	P33703 canis fami
36	216.5	14.5	646	1 LEM3_BOVIN	P42201 bos taurus
37	210	14.0	372	1 LEM1_MOUSE	P18337 mus musculu
38	203.5	13.6	372	1 LEM1_HUMAN	P14151 homo sapien
39	203.5	13.6	372	1 LEM1_PANTR	Q95237 pan troglod
40	202.5	13.5	372	1 LEM1_PONPY	Q95235 pongo pygma
41	200	13.4	372	1 LEM1_RAT	P30836 rattus norv
42	194	13.0	372	1 LEM1_MACMU	Q95196 macaca mula
43	194	13.0	372	1 LEM1_PAPHA	Q28768 papio hama
44	192.5	12.9	270	1 CFAH_HUMAN	P36980 homo sapien
45	192	12.8	297	1 APOH_RAT	P26644 rattus norv

## ALIGNMENTS

RESULT 1  
ID CFAH\_HUMAN STANDARD: PRT: 1231 AA.  
AC P08603;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HFI OR HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP TISSUE=LIVER;  
RC SEQUENCE FROM N.A.  
RX MEDLINE: 86134059.  
RA Ripchoe J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H.";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 53-445 FROM N.A.  
RX MEDLINE: 87054207.  
RA Schultz T.F., Schwaebler W., Stanley K.K., Weiss E., Dierich M.P.;  
RT "Human complement factor H: Isolation of cDNA clones and partial cDNA  
sequence of the 38-kDa tryptic fragment containing the binding site  
for C3b.";  
RL Eur. J. Immunol. 16:1351-1355(1986).  
RN [3]  
RP SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with C4b  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [4]  
RP SEQUENCE OF 1047-1231 FROM N.A.  
RX MEDLINE: 91201992.  
RA Estaller C., Kolstuen V., Schwaebler W., Dierich M.P., Weiss E.H.;  
RT "Cloning of the 1.4-kb mRNA species of human complement factor H  
reveals a novel member of the short consensus repeat family related  
to the carboxy terminal of the classical 150-kDa molecule.";  
RL J. Immunol. 146:3190-3196(1991).  
RN [5]  
RP SEQUENCE OF 19-35.  
RX MEDLINE: 83048213.  
RA Sim R.B., Discipio R.G.;  
RT "Purification and structural studies on the complement-system control  
protein beta 1H (factor H).";  
RL Biochem. J. 205:285-293(1982).  
RN [6]  
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).  
RX MEDLINE: 91278097.  
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim R., Campbell I.D.;  
RT "Three-dimensional structure of a complement control protein module  
in solution.";  
RL J. Mol. Biol. 219:717-725(1991).







```

FT SIGNAL 1 48
FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 2.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.
FT REPEAT 425 481 SUSHI 7.
FT REPEAT 483 539 SUSHI 8.
FT DISULFID 50 96 BY SIMILARITY.
FT DISULFID 81 108 BY SIMILARITY.
FT DISULFID 113 154 BY SIMILARITY.
FT DISULFID 140 170 BY SIMILARITY.
FT DISULFID 175 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 267 294 BY SIMILARITY.
FT DISULFID 299 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 365 387 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAc. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAc. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAc. . .).
FT VARIANT 92 92 Q -> T.
FT VARIANT 357 357 /FTIG-VAR_001977.
FT VARIANT 357 357 Y -> H.
FT VARIANT 357 357 /FTIG-VAR_001978.
SQ SEQUENCE 597 AA; 67033 MW; 67E03F2E8A5A16DD CRC64;

```

Query Match 20.2%; Score 303; DB 1; Length 597;  
 Best Local Similarity 30.7%; Pred. No. 1.5e-17;  
 Matches 80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;

```

QY 20 SDQTYEGTQAIYKCRPGY-RSLGNVIMWCR-GEWVALNPLKCKRPGCHGDTPECT 77
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 TETRFKGTTLKTYCLPGYVRSHTQTLGNSDGEWV-YWTF--CIYKRCRHGPE----- 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 FTLTGAVN-----FEYGVKAVYTCNGSYOLGELINRECDT---GWNDDIPICEVVC 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 --LRNGVEIKTDLSPGSOIEFSCSEGFLLIGTSTR-CEVODRGVSMHPLFOCEIVKC 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 LPYAPAREKIVSSAMPDREYHFGAVRFVCSGYKIEGDEEMHCSDD---GFWSKER 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 KPPEDIRNGHSGE---ENFYAVGVSIVTSCPPRSLIGHASISCTVEVETIGVWRPSP 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 PKCEVISCSPDYVINGSPISQ-KIYKENERFOYKCMGTEYSEKDAV--C-TESGMRP 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 PTCEKIKCRKRPDVSHGEMVSGFPIYNYKDTIVFKCKGKGVFL--KSSVITHCADSKMNP 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 L-PSCEKSCDN-pyIPNGDY 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 SPSPACEPNSCINLPDIPHASW 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 4
VCP_VACCV STANDARD; PRT; 263 AA.
ID VCP_VACCV
AC P10996;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)

```

```

DE (PROTEIN C3) (28 KDA PROTEIN).
GN C3.
OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
NC [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.
RC STRAIN=WR.
RX MEDLINE: 88318974.
RA Kotwal G.J., Moss B.;
RT "Vaccinia virus encodes a secretory polypeptide structurally related
  to complement control proteins.";
  Nature 335:176-178(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WR.
RX MEDLINE: 89073756.
RA Kotwal G.J., Moss B.;
RT "Analysis of a large cluster of nonessential genes deleted from a
  vaccinia virus terminal transposition mutant.";
  Virology 167:524-537(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COPENHAGEN;
RX MEDLINE: 91021027.
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
  Paolelli E.;
RT "The complete DNA sequence of vaccinia virus.";
  Virology 179:247-266(1990).
RN [4]
RP COMPLETE GENOME.
RC STRAIN=COPENHAGEN;
RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,
  Paolelli E.;
RT Virology 179:517-563(1990).
RN [5]
RP FUNCTION.
RX MEDLINE: 92115714.
RA Isaacs S.N., Kotwal G.J., Moss B.;
RT "Vaccinia virus complement-control protein prevents
  antibody-dependent complement-enhanced neutralization of infectivity
  and contributes to virulence.";
  Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).
RN [6]
RP STRUCTURE BY NMR OF 146-263.
RX MEDLINE: 97446168.
RA Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,
  Barlow P.N.;
RT "NMR studies of a viral protein that mimics the regulators of
  complement activation.";
  J. Mol. Biol. 272:253-265(1997).
RN [7]
RP FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY
  INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT
  ACTIVATION. BINDS C3B AND C4B.
RN [8]
RP SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF
  COMPLEMENT ACTIVATION (RCA).
RN [9]
RP SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.
RN [10]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X13166; CAA31564.1; -
CC EMBL: M22812; AAA69605.1; -
CC EMBL: M35027; AAA47997.1; -
CC PIR: A31005; MWVZSP.
CC PDB: 1VYC; 03-DEC-97.
CC PDB: 1VVD; 03-DEC-97.
CC PDB: 1VVE; 03-DEC-97.

```

DR INTERPRO: IPR000436; -  
 DR PFAM: PF00084; sushi; 4  
 KW Signal; Repeat; Sushi; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 263  
 FT DOMAIN 20 263  
 FT REPEAT 20 82  
 FT REPEAT 85 144  
 FT REPEAT 147 202  
 FT REPEAT 205 262  
 FT REPEAT 21 70  
 FT DISULFID 54 81  
 FT DISULFID 86 126  
 FT DISULFID 112 143  
 FT DISULFID 148 190  
 FT DISULFID 176 201  
 FT DISULFID 206 248  
 FT DISULFID 234 261  
 FT DISULFID 263 AA; 28629 MW; EA322C9A6F8997 CRC64;  
 S0 SEQUENCE

Query Match 20.1%; Score 301.5; DB 1; Length 263;  
 Best Local Similarity 33.3%; Pred. No. 7.9e-18;  
 Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;

QY 20 SDQYPEGQAIYKGRPGYR--SIGNVIMVCKGEMVALNPLKCKQKRCGHPGDTPTGT 77  
 Db 40 ANANYNIDDTLEYLCLPGRKOKMGPIYAKCTGTWLFN--QCIRKRCPSPRDIDMGQ 96  
 QY 78 FTLLGNAVFEYGVKAVYTCNKGYSLLIGET-NYRECDTGD---WTNDIPICEVVKCLPYTA 133  
 Db 97 LDIGG---VDFGSSITYSYCSNGYHLIGESKSYCELSTGSMWNNEAPKECVKQSPRS 153  
 QY 134 PENKIVSSAHEPREFHFGQAVRVCNSGYKIEDEMHCSDDGFWKSKPKVEISCK 193  
 Db 154 ISNGR--HNGYED--FYTDGSVYVYSCNSGYSLGNSGLCS-GGEWS-DPPTCQIVKCP 207  
 QY 194 SPDVINGSPISQ-KIYKENERFOYKCMNGYSEPDGAVCTESG-WRP-LPSC 244  
 Db 208 HPTISNGYLSGFPKRSYSYNVNDFKCKYKILSSSSSTCSFGNTWKPELPKC 261

RESULT 5  
 ID CRI\_HUMAN STANDARD; PRT; 2039 AA.  
 AC P17927;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN).  
 GN CRI OR C3BR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89035992.  
 RA Klickstein L.B., Barlow T.J., Miletic V., Rabson L.D., Smith J.A., Fearon D.T.;  
 RT "Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CRI, CD35) by deletion mutagenesis.";  
 RL J. Exp. Med. 168:1699-1717(1988).  
 RN [2]  
 RP SEQUENCE OF 503-2039 FROM N.A.  
 RX MEDLINE; 87168191.  
 RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G., Fearon D.T.;  
 RT "Human C3b/C4b receptor (CRI). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins.";  
 RL J. Exp. Med. 165:1095-1112(1987).  
 RN [3]

RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.  
 RX MEDLINE; 86067975.  
 RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;  
 RT "Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).  
 CC -I- FUNCTION: CRI, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE ACTIVATED COMPLEMENT.  
 CC -I- SUBUNIT: MONOMER.  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -I- POLYMORPHISM: CRI IS RESPONSIBLE FOR THE KNOBS BLOOD GROUP SYSTEM.  
 CC -I- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.  
 CC -I- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRI.  
 CC -I- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M11569; AAA52297.1; -  
 DR EMBL; M11617; AAA52298.1; -  
 DR EMBL; M11618; AAA52299.1; -  
 DR EMBL; Y00816; CAA68755.1; -  
 DR EMBL; X05309; CAA28933.1; -  
 DR PIR; A28507; A28507.  
 DR PIR; A24748; A24748.  
 DR PIR; B24748; B24748.  
 DR PIR; C24748; C24748.  
 DR PIR; S03843; S03843.  
 DR HSSP; P08603; 1HFI.  
 DR MIM; 120620; -  
 DR INTERPRO: IPR000436; -  
 DR PFAM: PF00084; sushi; 30.  
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;  
 KW Receptor; Sushi; Blood group antigen.  
 FT SIGNAL 1 41  
 FT CHAIN 2 2039  
 FT DOMAIN 42 1971  
 FT TRANSMEM 1972 1996  
 FT DOMAIN 1997 2039  
 FT MOD\_RES 42 42  
 FT DOMAIN 42 488  
 FT REPEAT 42 100  
 FT REPEAT 103 162  
 FT REPEAT 165 233  
 FT REPEAT 237 294  
 FT REPEAT 296 354  
 FT REPEAT 357 417  
 FT REPEAT 420 488  
 FT REPEAT 492 938  
 FT REPEAT 492 550  
 FT REPEAT 553 612  
 FT REPEAT 615 683  
 FT REPEAT 687 744  
 FT REPEAT 746 804  
 FT REPEAT 807 867  
 FT REPEAT 870 938  
 FT DOMAIN 942 1388  
 FT REPEAT 942 1000  
 FT REPEAT 1003 1062  
 CC COMPLEMENT RECEPTOR TYPE 1.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOLASMIC (POTENTIAL).  
 CC PYRROLIDONE CARBOXYLIC ACID (POTENTIAL).  
 CC 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS REPEAT A.  
 CC SUSHI A1.  
 CC SUSHI A2.  
 CC SUSHI A3.  
 CC SUSHI A4.  
 CC SUSHI A5.  
 CC SUSHI A6.  
 CC SUSHI A7.  
 CC 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS REPEAT B.  
 CC SUSHI B1.  
 CC SUSHI B2.  
 CC SUSHI B3.  
 CC SUSHI B4.  
 CC SUSHI B5.  
 CC SUSHI B6.  
 CC SUSHI B7.  
 CC 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS REPEAT C.  
 CC SUSHI C1.  
 CC SUSHI C2.

```

FT REPEAT 1065 1133 SUSHI C3.
FT REPEAT 1137 1194 SUSHI C4.
FT REPEAT 1196 1254 SUSHI C5.
FT REPEAT 1257 1317 SUSHI C6.
FT REPEAT 1320 1388 SUSHI C7.
FT DOMAIN 1395 1486 7 X SUSHI (SCR) REPEATS, LONG HOMOLOGOUS
REPEAT D.
FT REPEAT 1395 1453 SUSHI D1.
FT REPEAT 1456 1515 SUSHI D2.
FT REPEAT 1518 1586 SUSHI D3.
FT REPEAT 1590 1647 SUSHI D4.
FT REPEAT 1649 1707 SUSHI D5.
FT REPEAT 1710 1770 SUSHI D6.
FT REPEAT 1773 1841 SUSHI D7.
FT DOMAIN 1847 1966 2 X SUSHI (SCR) REPEATS.
REPEAT 1.
FT REPEAT 1908 1966 SUSHI 1.
FT DISULFID 43 86 BY SIMILARITY.
FT DISULFID 73 99 BY SIMILARITY.
FT DISULFID 104 145 BY SIMILARITY.
FT DISULFID 131 161 BY SIMILARITY.
FT DISULFID 166 215 BY SIMILARITY.
FT DISULFID 195 232 BY SIMILARITY.
FT DISULFID 238 280 BY SIMILARITY.
FT DISULFID 266 293 BY SIMILARITY.
FT DISULFID 297 340 BY SIMILARITY.
FT DISULFID 326 353 BY SIMILARITY.
FT DISULFID 358 400 BY SIMILARITY.
FT DISULFID 366 416 BY SIMILARITY.
FT DISULFID 421 470 BY SIMILARITY.
FT DISULFID 450 487 BY SIMILARITY.
FT DISULFID 493 536 BY SIMILARITY.
FT DISULFID 523 549 BY SIMILARITY.
FT DISULFID 554 595 BY SIMILARITY.
FT DISULFID 581 611 BY SIMILARITY.
FT DISULFID 616 665 BY SIMILARITY.
FT DISULFID 645 682 BY SIMILARITY.
FT DISULFID 688 730 BY SIMILARITY.
FT DISULFID 716 743 BY SIMILARITY.
FT DISULFID 747 790 BY SIMILARITY.
FT DISULFID 776 803 BY SIMILARITY.
FT DISULFID 808 850 BY SIMILARITY.
FT DISULFID 836 866 BY SIMILARITY.
FT DISULFID 871 920 BY SIMILARITY.
FT DISULFID 900 937 BY SIMILARITY.
FT DISULFID 943 986 BY SIMILARITY.
FT DISULFID 973 999 BY SIMILARITY.
FT DISULFID 1004 1045 BY SIMILARITY.
FT DISULFID 1031 1061 BY SIMILARITY.
FT DISULFID 1066 1115 BY SIMILARITY.
FT DISULFID 1095 1132 BY SIMILARITY.
FT DISULFID 1138 1180 BY SIMILARITY.
FT DISULFID 1166 1193 BY SIMILARITY.
FT DISULFID 1197 1240 BY SIMILARITY.
FT DISULFID 1226 1253 BY SIMILARITY.
FT DISULFID 1258 1300 BY SIMILARITY.
FT DISULFID 1286 1316 BY SIMILARITY.
FT DISULFID 1321 1370 BY SIMILARITY.
FT DISULFID 1350 1387 BY SIMILARITY.
FT DISULFID 1396 1439 BY SIMILARITY.
FT DISULFID 1426 1452 BY SIMILARITY.
FT DISULFID 1457 1498 BY SIMILARITY.
FT DISULFID 1484 1514 BY SIMILARITY.
FT DISULFID 1519 1568 BY SIMILARITY.
FT DISULFID 1548 1585 BY SIMILARITY.
FT DISULFID 1591 1633 BY SIMILARITY.
FT DISULFID 1619 1646 BY SIMILARITY.
FT DISULFID 1650 1693 BY SIMILARITY.
FT DISULFID 1679 1706 BY SIMILARITY.
FT DISULFID 1711 1753 BY SIMILARITY.
FT DISULFID 1731 1769 BY SIMILARITY.
FT DISULFID 1774 1823 BY SIMILARITY.
FT DISULFID 1803 1840 BY SIMILARITY.

```

```

FT DISULFID 1848 1891 BY SIMILARITY.
FT DISULFID 1877 1904 BY SIMILARITY.
FT DISULFID 1909 1952 BY SIMILARITY.
FT DISULFID 1938 1965 BY SIMILARITY.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 860 860 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 897 897 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 959 959 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1028 1028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1152 1152 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1310 1310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1481 1481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1540 1540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1605 1605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1763 1763 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 19.9%; Score 298.5; DB 1; Length 2039;
Best Local Similarity 27.0%; Pred. No. 1.4e-16;
Matches 86; Conservative 49; Mismatches 111; Indels 73; Caps 18;

OY 3 CNE--LPRRRTFELTGSWSQDOTPECTQAIYCRPGYSRLGNIVYCRKGEWVALNPL 59
   || || || || || || || || || || || || || || || || || || || ||
DB 43 CNAPEMLPAPARTNLT-----DEFEPITTYLNEBEPYSRGPSSILTKMSVTMGARD 97
   || || || || || || || || || || || || || || || || || || || ||
OY 60 KRCQKPCGHPEDPFGFTLTGCVNFEYGVKAVYTCNEGQLLGEINRYC-----DTDG 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 98 -RCRRKSCNPPDPVNGWVHYTK--IQGSGIKYSCRKGYRLGS--SSATCIISGDVI 153
   || || || || || || || || || || || || || || || || || || || ||
OY 116 WTNDDIPICEVVC-LPVTAPENGKIVSSAMEPDRYEHFGQAVRFVCSNG-----YKIEG 168
   || || || || || || || || || || || || || || || || || || || ||
DB 154 WNETPICRIKPCGLPPT-TNGDPISITNRE---NFHYGSVTTYICNCGSGRKVFELVG 209
   || || || || || || || || || || || || || || || || || || || ||
OY 169 DEEMHC--SDD--GFWSKKPKC-VEISCKSPDVINGSPIS--OKIIVKENERFOYKCNMG 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 210 EPSICTSMDQDVGLMSGAPQCIIIPNKCCTPPNVENGILVSDNRLFLNEVEFRCPG 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 223 YEYSEKGAIVCTE-SGWR-LPSCBE-----KSCDNP 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 270 FVMKGPBRVYCCALNKKWPELPSGRCVCOPPDVLAERTORDKDNFSPGQEVFYSCPG 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 253 Y-----IPNGDYSP 261
   | | | | | | | | | | | | | | | | | | | | |
DB 330 YDLRGAASMRCTPQGDMSF 348

RESULT 6
CABP RAT STANDARD; PRT; 558 AA.
ID CABP RAT
AC 063514;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUL-1998 (rel. 36, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.
GN C4BPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NC [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE; 97166082.
RA Hillarp A., Wiklund H., Thern A., Dahlback B.;
RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains:

```



RT	structural and functional relationships among human, bovine, rabbit,
RL	mouse, and rat proteins.";
J.	Immunol. 158:1315-1323(1997).
-I-	FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3bina), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A' ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERPIN ANGIOID P COMPONENT.
-I-	SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.
-I-	SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
-I-	SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement ( <a href="http://www.isb-sib.ch/announce/">see http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; 250051; CAA90391.1; .
DR	HSSP; P10998; IYVC.
DR	INTERPRO: IPR000436; .
DR	Pfam; PF00084; sushi; 8.
KW	Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
FT	SIGNAL 1 13 BY SIMILARITY.
FT	CHAIN 14 558 C4X-BINDING PROTEIN ALPHA CHAIN.
FT	DOMAIN 14 502 8 X SUSHI (SCR) REPEATS.
FT	REPEAT 14 73 SUSHI 1.
FT	REPEAT 76 135 SUSHI 2.
FT	REPEAT 138 200 SUSHI 3.
FT	REPEAT 203 259 SUSHI 4.
FT	REPEAT 262 325 SUSHI 5.
FT	REPEAT 328 387 SUSHI 6.
FT	REPEAT 389 444 SUSHI 7.
FT	REPEAT 446 502 SUSHI 8.
FT	DISULFID 15 60 BY SIMILARITY.
FT	DISULFID 45 72 BY SIMILARITY.
FT	DISULFID 77 118 BY SIMILARITY.
FT	DISULFID 104 134 BY SIMILARITY.
FT	DISULFID 139 182 BY SIMILARITY.
FT	DISULFID 168 199 BY SIMILARITY.
FT	DISULFID 204 246 BY SIMILARITY.
FT	DISULFID 232 258 BY SIMILARITY.
FT	DISULFID 263 312 BY SIMILARITY.
FT	DISULFID 296 324 BY SIMILARITY.
FT	DISULFID 7328 7351 BY SIMILARITY.
FT	DISULFID 7329 7373 BY SIMILARITY.
FT	DISULFID 7363 386 BY SIMILARITY.
FT	DISULFID 390 431 BY SIMILARITY.
FT	DISULFID 417 443 BY SIMILARITY.
FT	DISULFID 447 488 BY SIMILARITY.
FT	DISULFID 474 501 BY SIMILARITY.
FT	DISULFID 509 509 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	DISULFID 521 521 N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBOHYD 31 177 N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBOHYD 177 177 N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBOHYD 186 186 N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBOHYD 469 469 N-LINKED (GLNAC. . .) (POTENTIAL).
FT	CARBOHYD 491 491 N-LINKED (GLNAC. . .) (POTENTIAL).
SO	SEQUENCE 558 AA: 62266 MW: 592FC06V7EDIE5FF CRC64: 592FC06V7EDIE5FF

```

Query Match      18.5%  Score 276.5;  DB 1;  Length 558;
Best Local Similarity 30.2%  Pred. No. 2e-15;
Matches 83;  Conservative 43;  Mismatches 104;  Indels 45;  Gaps 17;

QY 7 PPRNTEILGSMWSDQT-YPECTQAIVKCRPPY-RSLGNVWCKR-GEW---VALNPLR 60
   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 17 PPDLPVALPASEMWQDTFESHITLRNCRPEYSSASSQSGLYCAFLGWQINIA----- 71

```

```

QY      61 KCKRPGHGADPFGFTLTGAVV-----FEYGVKAVYCNNGSYOLTG-ELIVRECDT 114
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      72 -CYKKSGRNGD-----LÖNGKVEKTKTFLFGSÖIEFSCSEDTYILIGSTSTCEIQK 123
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      115 G--WINDPICEVKK-LPYLADENKIVSSAMEPDRE---YHFGQAVRVCNSGYKIEG 168
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      124 GVSMSPDLPCVIYAKGMP---PD-----ISNGKHNGREEEFYRBSVYKCDPDLITLG 176
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      169 DEEMHDS---DGFMSKEKPKVEILSCSPDIVNSPIQ-KIYKENRFQYKCMNGY 223
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      177 NASITCTVYAKIVGWSPPPTCERIICWPYKVLHGHTNGFGFHTYKADSVRFQCKGF 236
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      224 EYSEKDAVCTESG-WRPLPSCSEKSK-DNPIYIPN 256
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      237 VLRSGVICHGADGWSMSPVPCVCELNSCTDIPDIPN 271
      11 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
C4BP_MOUSE
ID      C4BP_MOUSE      STANDARD;      PRT;      469 AA.
AC      P08607;
DT      01-AUG-1988 (rel. 08, Created)
DT      01-FEB-1994 (rel. 28, Last sequence update)
DT      01-NOV-1997 (rel. 35, Last annotation update)
DE      C4B-BINDING PROTEIN PRECURSOR (C4BP).
GN      C4BPA OR C4BP.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
RN      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      [1]
      SEQUENCE FROM N.A.
      MEDLINE: 88024997.
      Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
      "CDNA structure of murine C4b-binding protein, a regulatory component
      of the serum complement system.";
      Biochemistry 26:4668-4674(1987).
      -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
      ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
      (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
      ALSO ACCELERATES THE DEGRADATION OF THE C4C2A COMPLEX (C3
      CONVERTAS) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
      CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
      -!- SUBUNIT: HOMOPENTAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
      BETA CHAIN OF C4BP.
      -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
      -!- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.
      -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
      -----
      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
      or send an email to license@sib-sib.ch).
      -----
      CC      EMBL: M17122; AAA37312.1; ALT_INIT.
      DR      PIR: A27117; NBRM5C4.
      DR      HSSP; P10998; IVC.
      DR      MGD; MGI:88229; C4BP.
      DR      INTERPRO: IPR000436; -.
      KW      PFAM: PF00084; sushi; 6.
      DR      Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.
      FT      SIGNAL
      FT          1
      FT          56
      FT      CHAIN
      FT          57      469      C4B-BINDING PROTEIN.
      FT      DOMAIN
      FT          57      414      6 X SUSHI (SCR) REPEATS.
      FT      REPEAT
      FT          57      116      SUSHI 1.
      FT      REPEAT
      FT          119      177      SUSHI 2.
      FT      REPEAT
      FT          180      241      SUSHI 3.
      FT      REPEAT
      FT          244      300      SUSHI 4.
      FT      REPEAT
      FT          302      356      SUSHI 5.
      FT      REPEAT
      FT          358      414      SUSHI 6.
      FT      REPEAT
      FT          58      103      BY SIMILARITY.
      FT      DISULFID

```



FT DISULFID 88 115 BY SIMILARITY.  
 FT DISULFID 120 160 BY SIMILARITY.  
 FT DISULFID 146 176 BY SIMILARITY.  
 FT DISULFID 181 223 BY SIMILARITY.  
 FT DISULFID 209 240 BY SIMILARITY.  
 FT DISULFID 245 247 BY SIMILARITY.  
 FT DISULFID 273 299 BY SIMILARITY.  
 FT DISULFID 303 343 BY SIMILARITY.  
 FT DISULFID 329 355 BY SIMILARITY.  
 FT DISULFID 359 400 BY SIMILARITY.  
 FT DISULFID 386 413 BY SIMILARITY.  
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 469 AA; 5151 MM; 41E137CB8D8C6321 CRC64;

Query Match 18.4%; Score 275; DB 1; Length 469;  
 Best Local Similarity 28.8%; Pred. No. 2.2e-15;  
 Matches 78; Conservative 46; Mismatches 119; Indels 28; Gaps 15;

OY 7 PPRNTEILGMSDQT-YEGDQAIKCRPGY-RSLGNVIMCR-KGEM-VALNPLKRC 62  
 60 PPAIPALPASDVNRDFESHLLKYECLPGYGRGISRMVYCKPSGEWISVS---C 115  
 OY 63 QKRCRGHPGDPFGFTLTGNGVNEGVKAVYTCNEGTYOLLGEINTECDTDG---WTN 118  
 116 AKKCKRNGYIDNG---YVNGETTFSSQLEFSQDEFLVGS-STSCYKRGKGVAMSN 171  
 OY 119 DIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVREYNSGYKITEDEMHGS--- 175  
 172 PEPCEVIYKCGPPDISNGK--HSGTEDEFYXNMG--ISYTCDFGLVGSPEFICTVVN 227  
 OY 176 -DDGFWSKEKRYEISCKSPDVINGSPIS-QKITIKENEFQYCKMNGYISRGDANC 233  
 228 KTVPMWSSSPPTCKRIICSPNIIHGVISGYKATYTHRDSVRLACNGVYLRGRHVIEC 287  
 OY 234 TEGS-WRLPSCSEKSCD-NPIYPNGDYSPL 262  
 288 QGNGMWSLPTC-EFDCDLPAIYNGYITSM 317  
 Db

RESULT 8  
 MCP\_HUMAN STANDARD; PRT; 377 AA.  
 ID P15529;  
 AC 01-APR-1990 (Rel. 14, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST  
 DE LEPHOCYTE COMMON ANTIGEN) (TLX).  
 GN MCP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.  
 RX MEDLINE; 88286080.  
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,  
 RA Reberich M.B., Lemons R.S., Seya T., Atkinson J.P.,  
 RT "Molecular cloning and chromosomal localization of human membrane  
 RT cofactor protein (MCP). Evidence for inclusion in the multigene  
 RT family of complement-regulatory proteins."  
 RL J. Exp. Med. 168:181-194(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RX MEDLINE; 93119658.  
 RX Cervoni F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.;

RT "Characterization of a cDNA clone coding for human testis membrane  
 RT cofactor protein (MCP, CD46)."  
 RT Mol. Reprod. Dev. 34:107-113(1993).  
 RL [3]  
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE; 91267562.  
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,  
 RA McKenzie I.F.;  
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a  
 RT regulator of complement activation."  
 RT Immunogenetics 33:335-344(1991).  
 RL [4]  
 RN SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 94014356.  
 RA Cui W., Hourcade D., Post T., Greenland A.C., Atkinson J.P.,  
 RA Kumar V.;  
 RT "Characterization of the promoter region of the membrane cofactor  
 RT protein (CD46) gene of the human complement system and comparison to  
 RT a membrane cofactor protein-like genetic element."  
 RT J. Immunol. 151:4137-4146(1993).  
 RL [5]  
 RN ALTERNATIVE SPLICING.  
 RX MEDLINE; 92289809.  
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;  
 RT "Tissue-specific and allelic expression of the complement regulator  
 RT CD46 is controlled by alternative splicing."  
 RT Eur. J. Immunol. 22:1513-1518(1992)  
 CC -I- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST  
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY  
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD  
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3  
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH  
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT  
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE  
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -I- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE  
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN  
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE  
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND  
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F  
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME  
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I  
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS  
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICED EXON 13.  
 CC -I- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT  
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.  
 CC -I- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN  
 CC (PROBABLE).  
 CC -I- SIMILARITY: CONTAINS 4 SUSHT (SCR) REPEATS.  
 CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD46 entry;  
 CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD46.HTM"  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y00651; CAA68675.1; -  
 DR EMBL; S51940; AAB24802.1; -  
 DR EMBL; M58050; AAA62833.1; -  
 DR EMBL; A18585; CAA01400.1; -  
 DR EMBL; S65879; AAD13968.1; -  
 DR PIR; S01896; S01896.  
 DR HSSP; P10998; IYVC.  
 DR MIM; 120920; -  
 DR INTERPRO; IPR000436; -  
 DR PFAM; PF000084; susht; 4.  
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;

Sushi: Alternative splicing.

FT	SIGNAL	1	34	MEMBRANE COFACTOR PROTEIN.
FT	CHAIN	35	377	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	35	328	POTENTIAL.
FT	TRANSMEM	329	351	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	352	377	4 X SUSHI (SCR) REPEATS.
FT	REPEAT	35	284	SUSHI 1.
FT	REPEAT	35	95	SUSHI 2.
FT	REPEAT	161	158	SUSHI 3.
FT	REPEAT	227	224	SUSHI 4.
FT	REPEAT	227	284	SER/THR-RICH.
FT	DOMAIN	287	311	BY SIMILARITY.
FT	DISULFID	35	80	BY SIMILARITY.
FT	DISULFID	64	94	BY SIMILARITY.
FT	DISULFID	99	141	BY SIMILARITY.
FT	DISULFID	127	157	BY SIMILARITY.
FT	DISULFID	162	210	BY SIMILARITY.
FT	DISULFID	191	223	BY SIMILARITY.
FT	DISULFID	228	270	BY SIMILARITY.
FT	DISULFID	256	283	BY SIMILARITY.
FT	CARBOHYD	83	83	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	273	273	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	286	300	MISSING (IN A SECOND FORM).
FT	VARSPLIC	340	352	MISSING (IN ISOFORM M).
FT	VARSPLIC	353	361	YLORRKKG -> DIFKGRGRKQVVELMPLRLNQLQ
FT	VARSPLIC	362	377	OSRAE (IN ISOFORM M).
FT	VARSPLIC	362	377	TYLDETHREVEPTSL -> KADGAEVATYTKSTTAEQ
FT	VARSPLIC	362	377	RG (IN ISOFORM B/D/F/H/J/L).
FT	VARSPLIC	340	361	MISSING (IN ISOFORM M AND ISOFORM N).
FT	VARSPLIC	339	339	MISSING (IN ISOFORM N).
FT	VARSPLIC	339	339	I -> IGRQWELNMPRLRLNQLQOSRAE (IN ISOFORM N).
SO	SEQUENCE	377 AA:	42247 MW:	2CA6F61752570B57 CR64:

Query Match 18.3%; Score 274; DB 1; Length 377;  
 Best Local Similarity 29.5%; Pred. No. 2.1e-15;  
 Matches 77; Conservative 38; Mismatches 118; Indels 28; Gaps 13;

QY 2 DCNEPPRRNTEILGSGMSDQTPPGTQAIYKCRPGY--RELGNVIMCRGEGVALNP 58  
 Db 33 DACEPPPEFAMELIGK-PKPYTELGERVDYCKRGYIYPLATHTTIDRNMHTMLPVS 91  
 QY 59 LRCKGRPCGHDGDPFGFTLLTGNVFEYGVKAYTCNEGQQLG-EINRECDTG-- 115  
 Db 92 -DACYRETCPIYRD-PLNGQAVPANGTYEFQYGMHFCNEGYYLLIGELLY--CEIKGSV 147  
 QY 116 --WTDIDICEVVKCLPYTAPENKIVSSAMEPDEHYHFGQAVRFVCSNG----YKIEG 168  
 Db 148 AIWSGKPPICEVLCCTPPPKIKNGKHTSEVE--VFPEYLDVAIVYSCDPAPGDPFSLIG 204  
 QY 169 DEMHCSDDGFWSKPEKPCVEISCKSPDVINGSPIS--OKIITYENDEFOCKMNGYEV 225  
 Db 205 ESTIYCGNSWMSRAAPCKYKCFYVVENKQISGFKFYATYATWFE--CDKGYFL 262  
 QY 226 SERGDVAV-TESGW-RPLPSC 244  
 Db 263 DGSPTIYCDNSMTWDPVVKC 283

RESULT 9  
 LEM2\_HUMAN  
 ID LEM2\_HUMAN STANDARD: PRT; 610 AA.  
 AC P16581; P16111;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE E-SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)  
 DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)  
 DE (CD62E).  
 GN SELE OR ELAM1.  
 OS Homo sapiens (human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 90175359.  
 RA Hession C., Osborn L., Goff D., Chi-Rosso G., Vassallo C.,  
 RA Pasek M., Pittack C., Tizard R., Goetz S., McCarthy K., Hopple S.,  
 RA Lobb R.;  
 RT "Endothelial leukocyte adhesion molecule 1: direct expression cloning  
 RT and functional interactions.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1673-1677(1990).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 89162047.  
 RA Bevilacqua M.P., Stengelin S., Gimprone M.A. Jr., Seed B.;  
 RT "Endothelial leukocyte adhesion molecule 1: an inducible receptor for  
 RT neutrophils related to complement regulatory proteins and lectins.";  
 RL Science 243:1160-1165(1989).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE: 91115870.  
 RA Collins T., Williams A., Johnston G.L., Kim J., Eddy R., Shows T.,  
 RA Gimprone M.A. Jr., Bevilacqua M.P.;  
 RT "Structure and chromosomal location of the gene for endothelial-  
 RT leukocyte adhesion molecule 1.";  
 RL J. Biol. Chem. 266:2466-2473(1991).  
 RN [4]  
 RN LIGAND.  
 RX MEDLINE: 91068005.  
 RA Phillips M.L., Nudelma E., Gaeta F.C., Perez M., Singhal A.K.,  
 RA Hakomori S., Paulson J.C.;  
 RT "ELAM-1 mediates cell adhesion by recognition of a carbohydrate  
 RT ligand, sialyl-Lex.";  
 RL Science 250:1130-1132(1990).  
 RN [5]  
 RN 3D-STRUCTURE MODELING OF LECTIN DOMAIN.  
 RX MEDLINE: 93202275.  
 RA Mills A.;  
 RT "Modelling the carbohydrate recognition domain of human E-selectin.";  
 RL FEBS Lett. 319:5-11(1993).  
 RN [6]  
 RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 22-178.  
 RX MEDLINE: 94150646.  
 RA Graves B.J., Crowther R.L., Chandran C., Rumberger J.M., Li S.,  
 RA Huang K.-S., Presky D.H., Falleri P.C., Wolltzy B.A., Burns D.K.;  
 RT "Insight into E-selectin/ligand interaction from the crystal  
 RT structure and mutagenesis of the lec/ECF domains.";  
 RL Nature 367:532-538(1994).  
 RN [7]  
 RN VARIANT ARG-149.  
 RX MEDLINE: 95179107.  
 RA Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schatke S.,  
 RA Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;  
 RT "E-selectin polymorphism and atherosclerosis: an association study.";  
 RL Hum. Mol. Genet. 3:1935-1937(1994).  
 RN [8]  
 RN VARIANT ARG-149.  
 RX MEDLINE: 99134508.  
 RA Ye S.O., Usher D., Virgil D., Zhang L.Q., Yochim S.E., Gupta R.;  
 RT "A pect polymorphism detects the mutation of serine-128 to arginine in  
 RT CD 62E gene - a risk factor for coronary artery disease.";  
 RL J. Biomed. Sci. 6:18-21(1999).  
 CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND  
 CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY  
 CC POLYLACTOSAMINE X (ALPHA(1-3) FUCOSYLATED DERIVATIVES OF  
 CC GLYCOLIPIDS).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A  
 CC HIGHER RISK OF CORONARY ARTERY DISEASE (CAD). A SIGNIFICANTLY  
 CC HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH  
 CC ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN  
 CC UNSELECTED POPULATION (SER-149).



RT "Cloning and expression kinetics of porcine vascular cell adhesion molecule-1".

RL Biochem. Biophys. Res. Commun. 201:805-805(1994).

CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY ELAM-1 IS STAYL-L-LEWIS X (ALPHA(1->3)FUCCOSYLATED DERIVATIVES OF POLYACETOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF GLYCOPOLIPIDS).

CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT REJECTION AND PROBABLY ALSO IN XENOGRAFT REJECTION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMS.

CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS; PORCINE E-LECTIN LACKS THE HUMAN SUSHI-1 AND -4 EQUIVALENTS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U39076; AAA61545.1; -

DR EMBL: U08350; AAA21541.1; -

DR HSSP: P16581; 1EST.

DR INTERPRO: IPR000436; -

DR INTERPRO: IPR000561; -

DR INTERPRO: IPR001304; -

DR INTERPRO: IPR002396; -

DR PFAAM: PF000059; EGF\_1.

DR PFAAM: PF000059; lectin\_c; 1.

DR PRINTS: PR00343; SELECTIN.

DR PROSITE: PS00022; EGF\_1; 1.

DR PROSITE: PS01186; EGF\_2; 1.

DR PROSITE: PS00615; C-TYPE LECTIN\_1; 1.

DR PROSITE: PS00615; C-TYPE LECTIN\_2; 1.

DR Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;

KW Selectin; Signal; Sush1; Repeat.

FT SIGNAL 1 22

FT CHAIN 23 484

FT DOMAIN 23 429

FT TRANSMEM 430 451

FT DOMAIN 452 484

FT DOMAIN 39 139

FT DOMAIN 140 176

FT DOMAIN 180 421

FT REPEAT 180 236

FT REPEAT 239 299

FT REPEAT 302 362

FT REPEAT 365 421

FT REPEAT 41 139

FT DISULFID 112 131

FT DISULFID 144 155

FT DISULFID 149 164

FT DISULFID 166 175

FT DISULFID 181 222

FT DISULFID 208 235

FT DISULFID 240 285

FT DISULFID 271 298

FT DISULFID 303 348

FT DISULFID 334 361

FT DISULFID 366 407

FT DISULFID 393 420

FT CARBOHYD 61 61

FT CARBOHYD 65 65

FT CARBOHYD 79 79

FT CARBOHYD 160 160

FT CARBOHYD 201 201

FT CARBOHYD 254 254

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 253 253 C -> Y (IN REF. 2).

FT CONFLICT 313 313 L -> F (IN REF. 2).

FT CONFLICT 321 321 T -> N (IN REF. 2).

FT CONFLICT 327 327 K -> N (IN REF. 2).

FT CONFLICT 363 363 V -> A (IN REF. 2).

FT CONFLICT 384 384 V -> M (IN REF. 2).

FT CONFLICT 461 484 KFPSSSECTGPNRNGSYMSDIL -> NIFLPAPAFNFMPTKILIT (IN REF. 2).

FT SQ SEQUENCE 484 AA: 52567 MW: 48749.23 KDa

Query Match 17.8%; Score 266; DB 1; Length 484;

Best local Similarity 27.1%; Pred. No. 1.2e-14;

Matches 65; Conservative 42; Mismatches 103; Indels 30; Gaps 12;

QY 22 QTPREGQAIYKRCPRGSLGNVIMVC-RKGEVVALPLRCKCRPC--GHGDPFGF 77

DB 196 QSLPWNTTCAFECKEKEGELIGPEHLQCTSSGSDGKKP--TKAVTCDTVGHQPN--GD 250

QY 78 FTLTGCVFEYGVKAV--YTCNCGYOLGFIINRCEPTDG-WTNDIPICVVCPLPTAP 134

DB 251 VSCNHSIGGFAYKSTCHFTCAEGFGLQPAQI-ECTAQGQWTOQAFVCAVCAVPAVSOP 309

QY 135 ENKIVSSAMEPDREYHFGQAVRVCNSGYKIEGDEMHSCSDGFMSKEKPKVEISCKS 194

DB 310 KNG-LVFKTHSPTEFTYKSSCAFCSEGEFELKGSQQLACTSGQWMTQEPSPQVOCSS 368

QY 195 PDV-----INSGPISQKTIYKENERFOYKCMNGEYSRGAVCTESG-WR-PLPSC 245

DB 369 LEVPRELNSSGSEPVFVAVC-----TFACEGWMNLGNSVALTGATGCHMSMLPTCE 421

RESULT 11

C4BP\_BOVIN STANDARD; PRT; 610 AA.

AC 028065;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DI 01-NOV-1997 (Rel. 35, Last annotation update)

DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.

GN C4BPA.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

OC Bovidae; Bovinae; Bos.

RP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE: 95015909.

RA Hillarp A., Thern A., Dahlbaeck B.;

RT "Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains provides structural background for lack of complex formation with protein S."

RT J. Immunol. 153:4190-4199(1994).

RL J.

CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.

CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS (BY SIMILARITY).

CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.

CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC



```

CC or send an email to license@isb-sib.ch).
CC EMBL: M31516; AAA52169.1; -
DR EMBL: M30142; AAA52168.1; -
DR EMBL: M15799; AAA52167.1; -
DR EMBL: M64653; AAA52170.1; -
DR EMBL: M64356; AAA52170.1; JOINED.
DR EMBL: S72858; AAC60633.1; -
DR PIR: B26359; B26359.
DR PIR: A26359; A26359.
DR PIR: S16187; S16187.
DR PIR: A39101; A39101.
DR PIR: S23138; S23138.
DR HSSP: P08603; IHCC.
DR MIM: 125240; -.
DR INTERPRO: IPR000436; -.
DR PIR: PF00084; sushi; 4.
KW Complement pathway; Plasma; Glycoprotein; Membrane; Repeat;
KW Alternative splicing; GPI-anchor; Signal; Sushi; Polymorphism;
KW Blood group antigen.
FT SIGNAL 1 34
FT CHAIN 35 353 COMPLEMENT DECAT-ACCELERATING FACTOR.
FT PROPEP 354 381 REMOVED IN MATURE FORM.
FT DOMAIN 35 284 4 X SUSHI (SCR) REPEATS.
FT REPEAT 35 95 SUSHI 1.
FT REPEAT 97 159 SUSHI 2.
FT REPEAT 162 221 SUSHI 3.
FT REPEAT 224 284 SUSHI 4.
FT DOMAIN 287 356 SER/THR-RICH.
FT DISULFID 36 81
FT DISULFID 65 94
FT DISULFID 98 145
FT DISULFID 129 158
FT DISULFID 163 204
FT DISULFID 190 220
FT DISULFID 225 267
FT DISULFID 253 283
FT CARBOHYD 95 93
FT LIPID 353 353
FT VARSPIC 362 381
FT FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT FT GPI-ANCHOR.
FT FT HTCTLTGLGLTGLTGLTGLT -> SRPPTQAGRMCRSSLS
FT FT OSRTPEKRSFHSFLSPSSWYRAHVDFRFAWDASNHGIA
FT FT DIAKEELRRKYTOVRYLFLVS (IN ISOFORM DAF-1).
FT FT R -> L (IN TC(B) ANTIGEN).
FT FT /FTID-VAR_001997.
FT FT R -> P (IN TC(C) ANTIGEN).
FT FT /FTID-VAR_001998.
FT FT L -> R (IN MES(A) ANTIGEN).
FT FT /FTID-VAR_001999.
FT FT S -> L (IN DR(A-) ANTIGEN).
FT FT /FTID-VAR_002000.
FT FT A -> P (IN CR(A-) ANTIGEN).
FT FT /FTID-VAR_002001.
FT FT T -> I (IN REF. 1).
FT FT S -> M (IN REF. 2).
FT FT CONFLICT 80 80
FT FT CONFLICT 85 85
FT FT SEQUENCE 381 AA; 41388 MW; 29138EBBBA4B565E CRC64;
SQ

```

```

Query Match 17.5%; Score 262; DB 1; Length 381;
Best Local Similarity 29.8%; Pred. No. 2e-14;
Matches 71; Conservative 42; Mismatches 87; Indels 38; Gaps 12;

```

```

OY 1 EDCA---ELPRRNTTEITGSMOQY-PECTQALYKCRPYR---SLGNVIMCRKGEW 53
DB 92 EFCNCSCEVPTRLNSASLKOPYITQNYFPVGTVEYECRPREPSLSPKLTCLQMLKW 151
OY 54 VALNLRKCRKPGCHPDPTFGFTLTGNAVFEYGAAYTCNEGQQLLEIYVRECDT 113
DB 152 ---STAVECKKKKSCNPBEIRNGQIDVGGILF--GATISSCMTGKYLFEFSTS--SCLLI 206
OY 114 DG---WTNDIPICEVVCLEPVTAPENCKIVSSAMEPDRE-YHFGQAVFVCSGYKIEG 168
DB 207 SSSVQWMDPLPECREIYC-----PAPQIDNGIIGERDHYGRQSVTYACNKGFTMIG 261

```

```

OY 169 DEEMHC---SDDGFSKRPKC-----VEISCKSPDVIN-----GSPISOKIIVK 210
DB 262 EHSITCVNNDGEWSGPPCEKRSKLSITKVPPTVQKPTVNVPTTEVSPSQKTTTK 319
RESULT 14
ID LEM3_MOUSE STANDARD; PRT: 768 AA.
AC 001102;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE P-SELECTIN PRECURSOR (GRANULE MEMBRANE PROTEIN 140) (GMP-140) (PADGEM)
DE (CD62P) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 3) (LECAM3).
OS SELP OR GRMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE: 92340571.
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectins. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE: 92345617.
RA Sanders W.E. Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT "Molecular cloning and analysis of in vivo expression of murine P-
RT selectin.";
RL Blood 80:795-800(1992).
CC -I- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC THE CELL SURFACE.
CC -I- INDUCTION: BY TUMOR NECROSIS FACTOR ALPHA.
CC -I- SIMILARITY: TO OTHER SELECTIN/LECAMs.
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEATS; MOUSE P-LECTIN LACKS
CC THE HUMAN SUSHI-2 EQUIVALENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M87861; AAA40008.1; -
CC EMBL: M72332; AAA37712.1; -
CC DR PIR: A42755; A42755.
CC DR HSSP: P16109; IFSB.
CC DR MGD: MGI:98280; SELP.
CC DR INTERPRO: IPR000436; -.
CC DR INTERPRO: IPR000561; -.
CC DR INTERPRO: IPR001304; -.
CC DR INTERPRO: IPR002396; -.
CC DR PIR: PF00008; EGF; 1.
CC DR PIR: PF00059; lectin_c; 1.
CC DR PIR: PF00084; sushi; 8.
CC DR PRINTS: PR00343; SELECTIN.
CC DR PROSITE: PS00022; EGF_1; 1.
CC DR PROSITE: PS01186; EGF_2; 1.
CC DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
CC DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.

```



Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 KW Selectin; Signal; Sushi; Repeat; Palmitate.

FT CHAIN 1 41  
 FT SIGNAL 42 768  
 FT DOMAIN 42 709  
 FT TRANSMEM 710 733  
 FT DOMAIN 734 768  
 FT DOMAIN 158 195  
 FT DOMAIN 159 199  
 FT DOMAIN 199 700  
 FT REPEAT 159 700  
 FT REPEAT 199 700  
 FT REPEAT 258 282  
 FT REPEAT 261 320  
 FT REPEAT 323 382  
 FT REPEAT 385 444  
 FT REPEAT 447 506  
 FT REPEAT 509 568  
 FT REPEAT 579 638  
 FT REPEAT 641 700  
 FT DISULFID 60 158  
 FT DISULFID 131 150  
 FT DISULFID 163 174  
 FT DISULFID 168 183  
 FT DISULFID 185 194  
 FT DISULFID 200 244  
 FT DISULFID 230 257  
 FT DISULFID 262 306  
 FT DISULFID 292 319  
 FT DISULFID 324 368  
 FT DISULFID 354 381  
 FT DISULFID 386 430  
 FT DISULFID 416 442  
 FT DISULFID 448 493  
 FT DISULFID 478 505  
 FT DISULFID 510 554  
 FT DISULFID 540 567  
 FT DISULFID 580 624  
 FT DISULFID 610 637  
 FT DISULFID 642 686  
 FT DISULFID 672 699  
 FT CARBOHYD 398 398  
 FT CARBOHYD 603 603  
 FT CARBOHYD 654 654  
 FT CARBOHYD 661 661  
 FT CARBOHYD 679 679  
 FT LIPID 745 745  
 FT SITE 756 759  
 FT CONFLICT 724 724  
 FT SEQUENCE 768 AA: 83098 MW: E517307AD2FE6E68 CRC64;

Query Match 17.4%; Score 260.5; DB 1; Length 768;  
 Best Local Similarity 26.0%; Pred. No. 5.9e-14;  
 Matches 76; Conservative 47; Mismatches 106; Indels 63; Gaps 17;

QY 1 EDCELDPRRRTTEITGWSQD-----TYPEGTQAIYKCRPG--YRSLGNVIMCR 49  
 DB 129 QDCVEIYIKSNS--APGKWNDEPCFKRRALCYTASCDMSCSNGECIETIGSYTCGY 186  
 QY 50 K-----EWALNPLRKCK-----RPGHP-GDPTPTFTLLTGNNVFYVKAAY 94  
 DB 187 PGFGPDEEYV-----KCGGVNIPDHVLANCSHPLG-----FSFNSQCTF 228  
 QY 95 TCNNGYOL--LGEINRECDTDG--WTNDIPICEVVKCLPTVAPENGRKIVSSAMEPDRHYH 151  
 DB 229 SCAGYEYLDGEGEL---QCLASGIWNTNPPKCAVQCSGLAPRHGHT--ACMHPIAFA 283  
 QY 152 FGQAVRVNCSNGYKIEDDEHMGSDGDSWSEKRCVETISCKSPDV--INGG---PIQK 206  
 DB 284 YDSCKFECCPGYRARSNTLHCTGSGQWSEPLPTCEALICEPEIPITGHSMDCVPTGT 343  
 QY 207 IYKNERFQYKCMGYEYSERGDAVCTESG--W-RPLPSCEKSCDNVYIPN 256  
 DB 344 FGY--NSSCFELCAEGFVLGNDAIQCADSQWTAAPAFCEALQCPPEPVPS 393

RESULT 15  
 ID LEM2\_MOUSE STANDARD; PRT; 612 AA.  
 AC 000690;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-FEB-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE E-SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)  
 DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)  
 DE (CD62E)  
 GN SELE OR ELAM-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92283265.  
 RA Becker-Andre M., van Huijsduijn R.H., Losberger C., Whelan J.,  
 RA Delamarier J.F.;  
 RT "Murine endothelial leukocyte-adhesion molecule 1 is a close  
 RT structural and functional homologue of the human protein.";  
 RL Eur. J. Biochem. 206:401-411(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92340571.  
 RA Weller A., Isenmann S., Vestweber D.;  
 RT "Cloning of the mouse endothelial selectins. Expression of both E-  
 RT and P-selectin is inducible by tumor necrosis factor alpha.";  
 RL J. Biol. Chem. 267:15176-15183(1992).  
 CC -I- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND  
 CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY  
 CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF  
 CC POLYGLYCOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF  
 CC GLYCOLIPIDS)  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -I- SIMILARITY: TO OTHER SELECTINS/LECAMs.  
 CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC -I- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: M80778; AAA37547.1; -;  
 CC EMBL: M87862; AAA37577.1; ALT\_INIT.  
 CC DR HSSP: P16581; IKJA.  
 CC DR MGD: MGI:98278; SELE.  
 CC DR INTERPRO: IPR000436; -;  
 CC DR INTERPRO: IPR000561; -;  
 CC DR INTERPRO: IPR001304; -;  
 CC DR INTERPRO: IPR002396; -;  
 CC DR PFAM: PF00008; EGF\_1.  
 CC DR PFAM: PF00059; lectin\_c; 1.  
 CC DR PFAM: PF00084; sushi; 6.  
 CC DR PRINTS: PR00343; SELECTIN.  
 CC DR PRINTS: PS00022; BGF\_1; 1.  
 CC DR PROSITE: PS01186; BGF\_2; 1.  
 CC DR PROSITE: PS00615; C-TYPE\_LLECTIN\_1; 1.  
 CC DR PROSITE: PS50041; C-TYPE\_LLECTIN\_2; 1.  
 CC KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;  
 CC Selectin; Signal; Sushi; Repeat.  
 FT SIGNAL 1 21  
 FT CHAIN 22 612  
 FT DOMAIN 22 557  
 FT TRANSMEM 558 579  
 FT DOMAIN 580 612  
 CC E-SELECTIN.  
 CC EXTRACELLULAR (POTENTIAL).  
 CC POTENTIAL.  
 CC CYTOPLASMIC (POTENTIAL).







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:48 ; Search time 70.54 Seconds  
(without alignments)  
350.782 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497  
1 EDCNELPPRRNTETITGSMS.....EKSCDNPYIPNGDISPLAIK 265

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL.14:\*  
2: SP\_archaea:\*  
3: SP\_bacteria:\*  
4: SP\_fungi:\*  
5: SP\_human:\*  
6: SP\_invertebrate:\*  
7: SP\_mammal:\*  
8: SP\_mhbc:\*  
9: SP\_organelle:\*  
10: SP\_plant:\*  
11: SP\_protist:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1497	100.0	449	4	014570
2	699	46.7	669	6	028085
3	440	29.4	1053	13	091275
4	332.5	22.2	360	12	09YT08
5	322.5	21.5	550	12	P88903
6	322.5	21.5	550	12	040912
7	320	21.4	645	12	09WRU2
8	310	20.7	259	12	P87616
9	308.5	20.6	522	6	028769
10	301	20.1	679	11	099234
11	298.5	19.9	559	4	09UQV2
12	298.5	19.9	2039	4	016745
13	298.5	19.9	2489	4	016744
14	293	19.6	363	6	002839
15	290.5	19.4	315	6	028770
16	290.5	19.4	661	6	029531
17	289.5	19.3	497	6	029530
18	289.5	19.3	559	11	063612
19	289.5	19.3	559	11	063135

20	287	19.2	1911	6	029528	029528 papio hamad
21	284.5	19.0	263	12	089859	089859 variola vir
22	284.5	19.0	263	12	007033	007033 variola vir
23	283.5	18.9	263	12	089076	089076 variola vir
24	280.5	18.7	263	12	089061	089061 variola vir
25	279	18.6	1124	5	09YR4	09YR4 drosophila
26	277.5	18.5	560	5	022328	022328 caenorhabdi
27	276	18.4	522	6	028797	028797 pan troglod
28	274.5	18.3	349	4	015429	015429 homo sapien
29	272	18.2	285	6	019126	019126 macaca fasc
30	272	18.2	285	6	019127	019127 macaca mula
31	271	18.1	417	11	035520	035520 rattus norv
32	270.5	18.1	369	6	P79138	P79138 cercopithec
33	270.5	18.1	483	11	064735	064735 mus musculu
34	270.5	18.1	1652	5	09YIU9	09YIU9 drosophila
35	268.5	17.9	974	5	P91658	P91658 drosophila
36	267	17.8	285	6	019121	019121 papio hamad
37	266	17.8	482	6	028982	028982 sus scrofa
38	262	17.5	347	4	P78361	P78361 homo sapien
39	261	17.4	533	11	008569	008569 cavia porce
40	257	17.2	579	11	060736	060736 mus musculu
41	256.5	17.1	354	5	090611	090611 drosophila
42	256	17.1	657	4	014006	014006 homo sapien
43	255.5	17.1	1032	4	013866	013866 homo sapien
44	255	17.0	452	11	061407	061407 mus musculu
45	254.5	17.0	1045	6	046545	046545 ovis aries

## ALIGNMENTS

RESULT 1  
ID Q14570 PRELIMINARY: PRT: 449 AA.  
AC Q14570; P78435;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;  
RT "The complete amino acid sequence of human complement factor H.";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.;  
RT "Structural analysis of human complement protein H: homology with Cab  
binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RA Vik D.P., Williams S.A.;  
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Dominguez O.;  
RL Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.  
DR EMBL: X07523; CAAG0403.1; -  
DR EMBL: M12383; AAG52013.1; -  
DR EMBL: U56979; AAB01987.1; -  
DR EMBL: Z29665; CAA82763.1; -  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -  
DR PRAM: PF00084; sushi; 7.  
KW SIGNAL.  
FT SIGNAL 1 18 POTENTIAL.





[illegible]

Db	214	GEPSLYCTSKDDQYGIWISGAPQCIIIPNKKMPANVENCVLVSVRSLFSLNEVEFEPCOP	273
Qy	222	GVEYSEKGDVCTE-SCGRP-LPSCSEKSCDNPIYIPNDIYSP	261
Db	274	GFVMMGRPHRQCOALNWKPELPJSCSRVCOPPPEILHGHHP	315
RESULT	10		
ID	099254	PRELIMINARY;	PRT; 679 AA.
AC	099254;		
DT	01-NOV-1996 (TREMblrel, 01, Created)		
DT	01-NOV-1996 (TREMblrel, 01, Last sequence update)		
DT	01-MAY-2000 (TREMblrel, 13, Last annotation update)		
DE	COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).		
GN	CR2.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BALB/C;		
RX	MEDLINE: 90229754.		
RA	Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;		
RT	"The murine complement receptor gene family. IV. Alternative splicing		
RT	of Cr2 gene transcripts predicts two distinct gene products that share		
RL	homologous domains with both human CR2 and CR1."		
RN	J. Immunol. 144:3581-3591(1990).		
RP	(2)		
RP	SEQUENCE OF 21-367 FROM N.A.		
RX	MEDLINE: 95105691.		
RA	Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,		
RT	Holers V.M.;		
RT	"Mouse complement regulatory protein Cr1/p55 uses the specific		
RT	mechanisms of both human decay-accelerating factor and membrane		
RT	cofactor protein."		
RL	J. Exp. Med. 181:151-159(1995).		
DR	EMBL: U17128; AAA78271.1; JOINED.		
DR	EMBL: U17123; AAA78271.1; JOINED.		
DR	EMBL: U17124; AAA78271.1; JOINED.		
DR	EMBL: U17125; AAA78271.1; JOINED.		
DR	EMBL: U17126; AAA78271.1; JOINED.		
DR	EMBL: U17127; AAA78271.1; JOINED.		
DR	EMBL: M36470; AAA37449.1; -.		
DR	HSSP: P10998; IYVC.		
DR	MGD: MGI:88489; Cr2.		
DR	INTERPRO: IPR000436; -.		
DR	PFAM: PF00084; sushi; 10.		
FT	NON_TER 679		
FT	SEQUENCE 679 AA; 74916 MW; 52PC00FDCED20CDC CRC64;		
Query Match	20.1%;	Score 301;	DB 11; Length 679;
Best Local Similarity	26.0%;	Pred. No. 7.6e-20;	
Matches	82;	Conservative 51;	Mismatches 94; Indels 88; Gaps 16;
Qy	20	SDQT-VPEGTQAIYCRPQYRSLGNVIMYCRGEMVALNPLRCKQRP CGHGPDPFGTF	78
Db	37	SPKSEFAIGTWEYKCRPGYFRKSFILTCLETSKW--SDAQFCRKRKPPOMNPOEPLHGSV	94
Qy	79	TLTGCVFVEYGVKAVYTCNCGYGLGELINRREC-----PDDGWTNDIPITEVYKCLPYTPAP	134
Db	95	HINTG--IEFGSTIYSCNQGRLIGD--SSATCIYSDNTVMMDNDPLDESIPCESPPAI	151
Qy	135	ENGKIYSSAMEDDREHFCQAVRFVONSQ-----YKIGDEDMCHSD-----GFSKSK	184
Db	152	SNQDFSSSRD---SFFYGVAVTYTCHTKNREKFLDVLGERSIYCTSDNDQYGINSP	208
Qy	185	PKCV-EISCKSPDVING-----SPIS-----	204
Db	209	PCQIPRVKCPMPETIENGLESQFKHSFPLNDYIFKCKSGFTWKSRIAMQPNKRWSP	268
Qy	205	-----OKIIT-----KENERF-----QYKNMGVEYSEKGDVCTESG-W-RDL	241

Db	269	LPTCMGCLPQNTLHGKYNKKDEFFSVGOKVSYCNCNGYTLTGINTLVECTSLGTSMTV	328
OY	242	PSCEKSCDNPYTPN 256	
Db	329	PTCEVKSCD--AIPN 341	
RESULT	11		
O90QV2			
ID	090QV2	PRELIMINARY;	PRT; 559 AA.
AC	090QV2		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)		
DE	HDMAN CRI MRNA FOR C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 89010527.		
RA	Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;		
RT	"Identification of an alternative polyadenylation site in the human		
RT	C3b/C4b receptor (complement receptor type 1) transcriptional unit and		
RT	prediction of a secreted form of complement receptor type 1."		
RL	J. Exp. Med. 168:1255-1270(1988).		
DR	EMBL: X14362; CAA32541.1; -.		
DR	INTERPRO: IPR000436; -.		
DR	INTERPRO: IPR000834; -.		
DR	PFAM: PF000084; sushi; 8.		
DR	PROSITE: PS00133; CARBOXYPEPT_ZN2; UNKNOWN_1.		
KW	Signal.		
FT	NON_TER	1	1
FT	SIGNAL	<1	16
FT	CHAIN	17	559
SO	SEQUENCE	559 AA; 61424 MW; DBFFE965CA179D75 CRC64;	POTENTIAL. POTENTIAL.
Query Match		19.9%;	Score 298.5; DB 4; Length 559;
Best Local Similarity		27.0%;	Pred. No. 1e-19;
Matches	86; Conservative	49; Mismatches	111; Indels 73; Gaps 18.
OY	3	CNE---LPPRRNTEILTSGMSDQTYEGQAIYKCGPYRGYISGNTVMCGRGWVALNPL 59	
Db	18	CNAPEMLPPARPPNLT---DEFFPIGTYLNEYCRPGISGRPFSLICLKNSVYTGAKD- 72	
OY	60	RKCKRPCKHPGDTPGFTFLINGVFEYGAVALYVCNCGSYQLGEINYNREC---DTDG 115	
Db	73	-KRRKSCSNPPDPVAGMNVHVIKG--IQGSOIKKYSCTGRIIGS-SSATCIISSDYI 128	
OY	116	WTNDIPICEVVK-LFVTAPENGKIVSSAMEPDRYHFGQAVFVCNSG-----KIEG 168	
Db	129	WDNETPICRIPIGCLPPTI-TNGDFISTNRE--NPHYGSVYTYRCNPGSGGRKVELWG 184	
OY	169	DESMHC--SDD--GFPSKPKPK-VESICKSPVINGSPIS-QKIIYKEMEROYVCNNG 222	
Db	185	EPSTIYTSNDQYIGMSGAPQCIIPNCKTPAVNENGILLVSDNRSLFSLNEVVEFGQPG 244	
OY	223	YEYSEGDVACTE-SCMRP-LPSCCE-----KSCDNP 252	
Db	245	FVWKGRPRKCAQLNKEPELIPSCSKVYCCOPPDVLAERTQDKDNFSPQGEVYFSCERG 304	
OY	253	Y-----IPNGDYS 261	
Db	305	YDLRGAASMRCTPGDWSP 323	
RESULT	12		
ID	016745	PRELIMINARY;	PRT; 2039 AA.
AC	016745;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 RN Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94065175.  
 RA Vik D.P., Wong W.W.;  
 RT "Structure of the gene for the F allele of complement receptor type 1  
 and sequence of the coding region unique to the S allele."  
 J. Immunol. 151:6214-6224(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Vik D.P., Wong W.W.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L17418; AAB60694.1; -.  
 DR EMBL; L17390; AAB60694.1; JOINED.  
 DR EMBL; L17391; AAB60694.1; JOINED.  
 DR EMBL; L17392; AAB60694.1; JOINED.  
 DR EMBL; L17393; AAB60694.1; JOINED.  
 DR EMBL; L17394; AAB60694.1; JOINED.  
 DR EMBL; L17395; AAB60694.1; JOINED.  
 DR EMBL; L17396; AAB60694.1; JOINED.  
 DR EMBL; L17397; AAB60694.1; JOINED.  
 DR EMBL; L17398; AAB60694.1; JOINED.  
 DR EMBL; L17399; AAB60694.1; JOINED.  
 DR EMBL; L17400; AAB60694.1; JOINED.  
 DR EMBL; L17401; AAB60694.1; JOINED.  
 DR EMBL; L17402; AAB60694.1; JOINED.  
 DR EMBL; L17403; AAB60694.1; JOINED.  
 DR EMBL; L17404; AAB60694.1; JOINED.  
 DR EMBL; L17405; AAB60694.1; JOINED.  
 DR EMBL; L17406; AAB60694.1; JOINED.  
 DR EMBL; L17407; AAB60694.1; JOINED.  
 DR EMBL; L17408; AAB60694.1; JOINED.  
 DR EMBL; L17409; AAB60694.1; JOINED.  
 DR EMBL; L17410; AAB60694.1; JOINED.  
 DR EMBL; L17411; AAB60694.1; JOINED.  
 DR EMBL; L17412; AAB60694.1; JOINED.  
 DR EMBL; L17413; AAB60694.1; JOINED.  
 DR EMBL; L17414; AAB60694.1; JOINED.  
 DR EMBL; L17415; AAB60694.1; JOINED.  
 DR EMBL; L17416; AAB60694.1; JOINED.  
 DR EMBL; L17417; AAB60694.1; JOINED.  
 DR EMBL; L17419; AAB60694.1; JOINED.  
 DR EMBL; L17420; AAB60694.1; JOINED.  
 DR EMBL; L17421; AAB60694.1; JOINED.  
 DR EMBL; L17422; AAB60694.1; JOINED.  
 DR EMBL; L17423; AAB60694.1; JOINED.  
 DR HSP; P08603; IHL.  
 DR INTERPRO: IPR000436; -.  
 DR INTERPRO: IPR000834; -.  
 DR INTERPRO: IPR001424; -.  
 DR PFAW; PF000084; sushi; 30.  
 DR PROSITE; PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_2.  
 SQ SEQUENCE 2039 AA; 223603 MW; B82FCB11C6B16635 CRC64;

Query Match 19.9%; Score 298.5; DB 4; Length 2039;  
 Best Local Similarity 27.0%; Pred. No. 5e-19;  
 Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;

OY 3 CNE---LPPRRTTELITGMSQITPEGQAIYKCRPGTRSGNVIYVCRKGEWALNPL 59  
 DB 43 CNAPEWLPAPRTNLT---DEFEPIGTIYINYECPGSGRPFSTICLKNWVGAKD- 97  
 OY 60 RKCQKPCGHGDDPFGFTLTGAVFEYGAAYTCNGCYOLLGEINREC---DTGG 115  
 DB 98 -RCRRKSCRNPDPVNGAVHVIKQ--IQGSOIKISCTKGYRLTGS--SATCIIISGDTVI 153

OY 116 WTNDIPICEVVK-LPTAPENGKIVSSAMEPDEYHFGQAVREVCNSG-----YKIEG 168  
 DB 154 WDNETPICDRIPCLPPTI-TNGDFISTNRE---NFHYGSVVTYRCNNGSGGRKVFELVG 209  
 OY 169 DEEMHC--SDD--GEFSEKPKC-VFISCKSPDVINSPTS-QATIKENRRFYKCMG 222  
 DB 210 EPSTICTSNDQGVIGVSPAPQCTIPNKCCTPPNVENGILVSDNLSFLNEVFEEKCPG 269  
 OY 223 YEYSEKGDVCTE-SGNRP-LPSCSE-----KSCDNP 252  
 DB 270 FVMKGPFRVCOALNKNKEPELPSCSRVCQPPPDVIAHERRQKDNPSQGEVRYSCFPG 329  
 OY 253 Y-----IPNGDYSP 261  
 DB 330 YDLNGAASMRCTPGDWSP 348  
 RESULT 13  
 ID 016744 PRELIMINARY; PRT; 2489 AA.  
 AC 016744;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 RN Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94065175.  
 RA Vik D.P., Wong W.W.;  
 RT "Structure of the gene for the F allele of complement receptor type 1  
 and sequence of the coding region unique to the S allele."  
 J. Immunol. 151:6214-6224(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Vik D.P., Wong W.W.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L17418; AAB60695.1; -.  
 DR EMBL; L17390; AAB60695.1; JOINED.  
 DR EMBL; L17391; AAB60695.1; JOINED.  
 DR EMBL; L17392; AAB60695.1; JOINED.  
 DR EMBL; L17393; AAB60695.1; JOINED.  
 DR EMBL; L17394; AAB60695.1; JOINED.  
 DR EMBL; L17395; AAB60695.1; JOINED.  
 DR EMBL; L17396; AAB60695.1; JOINED.  
 DR EMBL; L17397; AAB60695.1; JOINED.  
 DR EMBL; L17398; AAB60695.1; JOINED.  
 DR EMBL; L17399; AAB60695.1; JOINED.  
 DR EMBL; L17400; AAB60695.1; JOINED.  
 DR EMBL; L17401; AAB60695.1; JOINED.  
 DR EMBL; L17402; AAB60695.1; JOINED.  
 DR EMBL; L17403; AAB60695.1; JOINED.  
 DR EMBL; L17404; AAB60695.1; JOINED.  
 DR EMBL; L17405; AAB60695.1; JOINED.  
 DR EMBL; L17406; AAB60695.1; JOINED.  
 DR EMBL; L17407; AAB60695.1; JOINED.  
 DR EMBL; L17408; AAB60695.1; JOINED.  
 DR EMBL; L17409; AAB60695.1; JOINED.  
 DR EMBL; L17410; AAB60695.1; JOINED.  
 DR EMBL; L17411; AAB60695.1; JOINED.  
 DR EMBL; L17412; AAB60695.1; JOINED.  
 DR EMBL; L17413; AAB60695.1; JOINED.  
 DR EMBL; L17414; AAB60695.1; JOINED.  
 DR EMBL; L17415; AAB60695.1; JOINED.  
 DR EMBL; L17416; AAB60695.1; JOINED.  
 DR EMBL; L17417; AAB60695.1; JOINED.  
 DR EMBL; L17419; AAB60695.1; JOINED.  
 DR EMBL; L17420; AAB60695.1; JOINED.  
 DR EMBL; L17421; AAB60695.1; JOINED.  
 DR EMBL; L17422; AAB60695.1; JOINED.  
 DR EMBL; L17423; AAB60695.1; JOINED.  
 DR EMBL; L17424; AAB60695.1; JOINED.



Query Match 19.68; Score 293; DB 6; Length 363;

[illegible]

QY 255 -----PNGDYSP 261  
| | | |  
Db 287 PSYDLRGASLRCTPOGDWSP 307

Search completed: November 21, 2000, 16:48:51  
Job time: 329 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:39:01 Search time 49.92 Seconds  
(without alignments)  
141.789 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163  
Sequence: 1 EDCEMLPPRRNTETITGWSWS.....VEISCKSPDVINGSPISQKT 207

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72.0	62.0	240	18 W39154	Human partial Comp
2	62.7	53.9	216	18 W39155	Clome PRRR9FH410 C
3	62.7	53.9	216	18 W39155	Decay accelerating
4	62.7	53.9	216	18 W39155	Decay accelerating
5	62.7	53.9	216	18 W39155	Human complement r
6	62.7	53.9	216	18 W39155	Human decay accele
7	62.7	53.9	216	18 W39155	Decay accelerating
8	62.7	53.9	216	18 W39155	Decay accelerating
9	62.7	53.9	216	18 W39155	Human decay accele
10	62.7	53.9	216	18 W39155	Human CD55 and 791
11	62.7	53.9	216	18 W39155	Decay accelerating
12	62.7	53.9	216	18 W39155	Human decay accele
					Membrane bound dec

13	261	22.4	440	16 R65684	Decay accelerating
14	261	22.4	440	18 W27483	Human glycopospha
15	261	22.4	543	13 R28557	CR1-4 (99H, 103E)
16	261	22.4	577	17 W06882	Membrane co-factor
17	259	22.3	543	13 R28547	CR1-4 (52S, 53S, 5
18	257.5	22.1	496	20 Y55752	Human CR1 protein
19	257	22.1	543	13 R28560	CR1-4 (114S) analo
20	255	21.9	254	15 R47154	Sequence of solubl
21	255	21.9	254	15 R47155	CR1-4 (64K) analog
22	255	21.9	543	13 R28550	CR1-4 (85R, 87N) a
23	255	21.9	543	13 R28553	CR1-4 (121Q) analo
24	255	21.9	543	13 R28565	CR1-4 (318R, 319N)
25	255	21.9	543	13 R28566	CR1-4 (318-321 RNP
26	255	21.9	543	13 R28567	CR1-4 (347T, 349Y)
27	255	21.9	543	13 R28568	CR1-4 (369-376 STR
28	255	21.9	543	13 R28569	CR1-4 (266-274 KIK
29	255	21.9	543	13 R28570	CR1-4 (364-367 NAA
30	255	21.9	543	13 R28571	Amino acid sequenc
31	255	21.9	1930	19 W73147	Human complement r
32	255	21.9	2039	12 R11810	Human C3b/C4b rece
33	255	21.9	2039	12 R11810	Human C3b/C4b rece
34	255	21.9	2039	12 R11810	Human C3b/C4b rece
35	255	21.9	2039	12 R11810	Human C3b/C4b rece
36	254	21.8	543	13 R28545	CR1-4 (57V, 59K) a
37	254	21.8	543	13 R28548	CR1-4 (57V, 59K) a
38	253.5	21.8	484	21 Y81934	Porcine E-selectin
39	253.5	21.8	484	21 Y81934	Porcine E-selectin
40	253	21.8	543	13 R28555	CR1-4 (109N, 110A)
41	253	21.8	543	13 R28558	CR1-4 (117P) analo
42	253	21.8	543	13 R28563	CR1-4 (64K, 65T)
43	252	21.7	543	13 R28549	CR1-4 (65T) analog
44	252	21.7	543	13 R28549	CR1-4 (65T) analog
45	252	21.7	543	13 R28562	CR1-4 (116K) analo

# ALIGNMENTS

RESULT 1	
W39154	W39154 standard; Protein: 240 AA.
AC	W39154:
XX	
XX	27-APR-1998 (first entry)
DT	
DE	Human partial Complement factor H protein fragment 1.
XX	
XX	Complement factor H; tumour associated antigen; renal cancer;
KW	urogenital cancer; medicament; modulator.
KW	
XX	
OS	Homo sapiens.
XX	
XX	W09738136-A1.
PN	
XX	16-OCT-1997.
PD	
XX	
PF	09-APR-1997; 97WO-US05710.
XX	
PR	06-MAR-1997; 97US-0812481.
XX	
PR	09-APR-1996; 96US-0015083.
XX	
PR	09-APR-1996; 96US-0630048.
XX	
PR	06-MAR-1997; 97US-0038614.
XX	
PA	(BARD-) BARD DIAGNOSTIC SCI INC.
XX	
PI	Enfield DL, Hass GM, Kinders RJ;
XX	
DR	WPI; 1997-512742/47.
XX	
DR	N-PSDB; V02790.
XX	
PT	Treating or screening for cancer, e.g. renal or urogenital cancer -
PT	by modulating or detecting tumour associated human complement factor

PT H related antigen, or nucleic acid encoding it  
 XX  
 XX Example 6B; Fig 6B; 104pp; English.  
 CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone pRAB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigens can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.  
 CC  
 SQ Sequence 240 AA;  
 QY  
 Query Match 62.0%; Score 721; DB 18; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-59;  
 Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 1 fcltgnvteygykavvtcmegyllgelnytcddtgwctndipicevkcclptapeng 60  
 QY 78 FTLTGQNVFEYGYKAVYTCNEGYOLLGEINVRCDTGDGTNDIPICGVKCLPTAPENG 137  
 Db 1 fcltgnvteygykavvtcmegyllgelnytcddtgwctndipicevkcclptapeng 60  
 QY 138 KIVSSAMEPDRREYHFGQAVFVNCNSGYKTECDDEMHCSDDGFWSKKEPKCVETSCSKSPDY 197  
 Db 61 klvssamepdrrehyhfgqavrfvncnsgykiedgmhcsddgfwskepkcvetlsckspdy 120  
 QY 198 INGSPISSQKI 207  
 Db 121 lngspisqkl 130  
 Db  
 RESULT 2  
 W39155  
 ID W39155 standard; Protein: 216 AA.  
 AC W39155;  
 XX  
 AC W39155;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Clone pRAB9FH410 CFH related protein fragment.  
 XX  
 KW Complement factor H; tumour associated antigen; renal cancer;  
 KW urogenital cancer; medicament; modulator.  
 OS Synthetic.  
 OS  
 PN W09738136-A1.  
 PD 16-OCT-1997.  
 PD  
 PF 09-APR-1997; 97WO-US05710.  
 PF  
 PR 06-MAR-1997; 97US-0812481.  
 PR 09-APR-1996; 96US-0015083.  
 PR 09-APR-1996; 96US-0630048.  
 PR 06-MAR-1997; 97US-0038614.  
 PR  
 XX  
 PA (BARD-) BARD DIAGNOSTIC SCI INC.  
 XX  
 PI Entfield DL, Hass GM, Kinders RJ;  
 XX  
 DR WPI; 1997-512742/47.  
 DR N-PSDB; V02791.  
 DR  
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
 PT by modulating or detecting tumour associated human complement factor  
 PT H related antigen, or nucleic acid encoding it  
 XX  
 XX Example 6B; Fig 6B; 104pp; English.  
 CC This partial protein is found in clone pRAB9FH410 and represents a

CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.  
 CC  
 SQ Sequence 216 AA;  
 QY  
 Query Match 53.9%; Score 627; DB 18; Length 216;  
 Best Local Similarity 99.1%; Pred. No. 6e-51;  
 Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Db 1 tcnegvllgeinvrccdtgwtndipicgvkclptapengkiyssamepdrrehyf 60  
 QY 155 AVRFVNCNSGYKTEGDEMHCSDDGFWSKKEPKCVETSCSKSPDYINSPISSQKI 207  
 Db 61 avrfvncnsgykiedgmhcsddgfwskepkcvetlsckspdyinspsqkl 113  
 Db  
 RESULT 3  
 W06881  
 ID W06881 standard; Protein: 299 AA.  
 AC W06881;  
 XX  
 AC W06881;  
 XX  
 DT 18-MAR-1997 (first entry)  
 XX  
 DE Decay accelerating factor.  
 XX  
 KW Complement inhibitor; membrane co-factor protein; MCP;  
 KW decay accelerating factor; DAF; chimeric protein; glycosaminoglycan;  
 KW heparin; cell lysis; sepsis; adult respiratory distress syndrome;  
 KW reperfusion injury; cell damage.  
 XX  
 OS Homo sapiens.  
 OS  
 PN W09634965-A2.  
 PD 07-NOV-1996.  
 PD  
 PF 03-MAY-1996; 96WO-US06301.  
 PF  
 PR 05-MAY-1995; 95US-0435149.  
 PR  
 XX  
 PA (CHIR) CHIRON CORP.  
 XX  
 PI Creasey AA, Innis MA, Zaror I;  
 XX  
 DR WPI; 1996-506167/50.  
 DR  
 PT Chimeric proteins for inhibiting complement-mediated cell lysis -  
 PT comprise membrane co-factor protein and decay accelerating factor  
 PT peptide sequences  
 XX  
 PS Disclosure: Page 25; 33pp; English.  
 XX  
 XX A portion (W06881) of the complement-inhibitor, decay accelerating  
 CC factor (DAF), is used in novel chimeric proteins of formula  
 CC A-R1-B-R2-C, where A and C are peptides (W06875-79, W06883-90) able  
 CC to bind glycosaminoglycans (esp. heparin) present on cell surfaces,  
 CC R1 is a portion of DAF or membrane co-factor protein (MCP, see also  
 CC W06880), R2 is DAF when R1 is MCP or MCP when R1 is DAF, and B is a  
 CC peptide that may have complement inhibitor activity. The chimeric  
 CC proteins (see also W06882) are directed to cell surfaces where they  
 CC inhibit complement-mediated cell lysis. They are used to treat and  
 CC prevent disease states in which complement plays a role, e.g.  
 CC sepsis, adult respiratory distress syndrome, reperfusion injury and  
 CC tissue damage.  
 CC



SQ Sequence 381 AA;  
 Query Match 22.4%; Score 261; DB 8; Length 381;  
 Best Local Similarity 29.9%; Pred. No. 1,4e-16;  
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;  
 QY 1 EDCN---ELPPRNTTEILTGSWSDQY-PECTQAIYCRPGYR---SIGNVIMYCRGEM 53  
 DB 92 efmrseceptrlnasalkpdytlqnyfpvgltvveyecrpyrrrepslptcltqlnkw 151  
 QY 54 VALNPLRCKCRKPCGHPDPEFTFTLLGQVVEYGVKAVYTCNKGQYQLGELINRECDT 113  
 DB 152 --stavefcckkscpnpgelnqgidvpgllf--gatisfscntgylkfgsts-sfcll 206  
 QY 114 DG---WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEG 168  
 DB 207 sgssvqwsdplpecrelyc-----pappqldngllqgerdhyyrgsvtyacnkqftmfg 261  
 QY 169 DEEMHC---SDDGFWSEKPKC-----VEISCKSPDVIN-----GSPISOK 206  
 DB 262 ehstlyctvmdegewsgpppecrgklskvpplvtqkptltvnpptleavpslsqk 315  
 RESULT 6  
 ID P94773 standard; protein; 381 AA.  
 AC P94773;  
 XX  
 XX  
 DT 04-JUL-1990 (first entry)  
 DE Decay accelerating factor (DAF) of clones lambda 33 and lambda 47.  
 XX  
 XX DAF; allograft rejection; affinity purification;  
 KW autoimmune disease; ds.  
 XX  
 XX Synthetic.  
 OS  
 XX  
 XX W08901041-A.  
 XX  
 XX 09-FEB-1989.  
 PD  
 XX  
 XX 03-AUG-1988; 88WO-US02648.  
 PF  
 XX 06-AUG-1987; 87US-0083757.  
 PR  
 XX (GETH ) GENETECH INC.  
 PA  
 XX Caras I;  
 PI  
 XX  
 XX WPI: 1989-061177/08.  
 DR N-PSDB; N91043.  
 DR  
 XX Fusion polypeptide for targeting protein to cell membrane -  
 PT comprises phospholipid anchor domain with heterologous  
 PT polypeptide.  
 PT  
 XX Disclosure; : 61pp; English.  
 PS  
 XX Recombinant DAF's are useful in treatment of inflammatory or cell lytic  
 CC autoimmune diseases and allograft rejection. Useful in diagnostic  
 CC compositions or in affinity purification.  
 CC  
 XX Sequence 381 AA;  
 SO  
 Query Match 22.4%; Score 261; DB 10; Length 381;  
 Best Local Similarity 29.9%; Pred. No. 1,4e-16;  
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;  
 QY 1 EDCN---ELPPRNTTEILTGSWSDQY-PECTQAIYCRPGYR---SIGNVIMYCRGEM 53  
 DB 92 efmrseceptrlnasalkpdytlqnyfpvgltvveyecrpyrrrepslptcltqlnkw 151  
 QY 54 VALNPLRCKCRKPCGHPDPEFTFTLLGQVVEYGVKAVYTCNKGQYQLGELINRECDT 113  
 DB 152 --stavefcckkscpnpgelnqgidvpgllf--gatisfscntgylkfgsts-sfcll 206  
 QY 114 DG---WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEG 168  
 DB 207 sgssvqwsdplpecrelyc-----pappqldngllqgerdhyyrgsvtyacnkqftmfg 261  
 QY 169 DEEMHC---SDDGFWSEKPKC-----VEISCKSPDVIN-----GSPISOK 206  
 DB 262 ehstlyctvmdegewsgpppecrgklskvpplvtqkptltvnpptleavpslsqk 315  
 RESULT 7  
 ID R66683 standard; protein; 381 AA.  
 AC R66683;  
 XX  
 XX  
 DT 23-JUL-1995 (first entry)  
 DE Decay accelerating factor.  
 XX  
 XX  
 DE Decay accelerating factor.  
 XX  
 XX  
 KW Decay accelerating factor; DAF; mDAF; fusion protein; liposome;  
 KW cell targeting; glycoposphatidylinositol; GPI; drug delivery.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /label= sig\_peptide  
 FT Modified-site 364  
 FT /note= "probable phosphatidylinositol  
 FT derivatization site"  
 XX  
 XX US5374548-A.  
 XX  
 XX 20-DEC-1994.  
 PD  
 XX  
 XX 02-MAY-1986; 86US-0859107.  
 PF  
 XX  
 XX 02-MAY-1986; 86US-0859107.  
 PR  
 XX 02-MAY-1986; 86US-0859107.  
 PR 06-AUG-1987; 87US-0083757.  
 PR 19-DEC-1991; 91US-0811048.  
 PR 12-FEB-1993; 93US-0017934.  
 PA  
 XX (GETH ) GENETECH INC.  
 PA  
 XX Caras IW;  
 PI  
 XX  
 XX WPI: 1995-035649/05.  
 DR N-PSDB; Q79863.  
 DR  
 XX Liposome(s) for targeting particular cells contg. fusion protein  
 PT - of glyco:phosphatidylinositol anchor and heterologous  
 PT targeting protein, e.g. for delivering toxins to infected or  
 PT cancer cells.  
 PT  
 XX Disclosure; Fig. 1a-1f; 36pp; English.  
 PS  
 XX A probe (given in Q79863) based on the N-terminal sequence of human  
 CC decay accelerating factor (DAF) was used to screen a HeLa cell  
 CC lambda cDNA library. Isolated clones encoding membrane-bound DAF  
 CC (mDAF) were obtained, and the full sequence of mDAF cDNA was  
 CC determined (Q79863). The GPI signal domain of mDAF may be fused to  
 CC a heterologous protein and targeted to cell membrane surfaces.  
 CC  
 XX Sequence 381 AA;  
 SO  
 Query Match 22.4%; Score 261; DB 16; Length 381;  
 Best Local Similarity 29.9%; Pred. No. 1,4e-16;  
 Matches 70; Conservative 42; Mismatches 84; Indels 38; Gaps 12;  
 QY 1 EDCN---ELPPRNTTEILTGSWSDQY-PECTQAIYCRPGYR---SIGNVIMYCRGEM 53  
 DB 92 efmrseceptrlnasalkpdytlqnyfpvgltvveyecrpyrrrepslptcltqlnkw 151  
 QY 54 VALNPLRCKCRKPCGHPDPEFTFTLLGQVVEYGVKAVYTCNKGQYQLGELINRECDT 113  
 DB 152 --stavefcckkscpnpgelnqgidvpgllf--gatisfscntgylkfgsts-sfcll 206  
 QY 114 DG---WTNDIPICEVVKCLPVTAPENKIVSSAMEPDRE-YHFGQAVRFVCSNGYKIEG 168  
 DB 207 sgssvqwsdplpecrelyc-----pappqldngllqgerdhyyrgsvtyacnkqftmfg 261  
 QY 169 DEEMHC---SDDGFWSEKPKC-----VEISCKSPDVIN-----GSPISOK 206  
 DB 262 ehstlyctvmdegewsgpppecrgklskvpplvtqkptltvnpptleavpslsqk 315  
 RESULT 7  
 ID R66683 standard; protein; 381 AA.  
 AC R66683;  
 XX  
 XX  
 DT 23-JUL-1995 (first entry)  
 DE Decay accelerating factor.  
 XX  
 XX  
 DE Decay accelerating factor.  
 XX  
 XX  
 KW Decay accelerating factor; DAF; mDAF; fusion protein; liposome;  
 KW cell targeting; glycoposphatidylinositol; GPI; drug delivery.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..34  
 FT /label= sig\_peptide  
 FT Modified-site 364  
 FT /note= "probable phosphatidylinositol  
 FT derivatization site"  
 XX  
 XX US5374548-A.  
 XX  
 XX 20-DEC-1994.  
 PD  
 XX  
 XX 02-MAY-1986; 86US-0859107.  
 PF  
 XX  
 XX 02-MAY-1986; 86US-0859107.  
 PR  
 XX 02-MAY-1986; 86US-0859107.  
 PR 06-AUG-1987; 87US-0083757.  
 PR 19-DEC-1991; 91US-0811048.  
 PR 12-FEB-1993; 93US-0017934.  
 PA  
 XX (GETH ) GENETECH INC.  
 PA  
 XX Caras IW;  
 PI  
 XX  
 XX WPI: 1995-035649/05.  
 DR N-PSDB; Q79863.  
 DR  
 XX Liposome(s) for targeting particular cells contg. fusion protein  
 PT - of glyco:phosphatidylinositol anchor and heterologous  
 PT targeting protein, e.g. for delivering toxins to infected or  
 PT cancer cells.  
 PT  
 XX Disclosure; Fig. 1a-1f; 36pp; English.  
 PS  
 XX A probe (given in Q79863) based on the N-terminal sequence of human  
 CC decay accelerating factor (DAF) was used to screen a HeLa cell  
 CC lambda cDNA library. Isolated clones encoding membrane-bound DAF  
 CC (mDAF) were obtained, and the full sequence of mDAF cDNA was  
 CC determined (Q79863). The GPI signal domain of mDAF may be fused to  
 CC a heterologous protein and targeted to cell membrane surfaces.  
 CC  
 XX Sequence 381 AA;  
 SO

	Best Local Similarity	29.9%, Pred. No. 1,4e-16;	
Matches	70; Conservative	42; Mismatches	84; Indels 38; Gaps 12;
QY	1 EDCN---ELPPRRNEILTGSWSDOTY--PEGTQAIVKCRPGR---SLGAVIWCCKGEM	53	
Db	92 efncrscvptrlnasalkqpyltgnyfpvqltveyeceprpgyirrepalspkllcqlnlfkw	151	
OY	54 VALNPRLRKOKRRCGPGDPTFGTELLGVWFEEGVAAVTCNEGQLGEINRYECRT	113	
Db	152 --stavefckkskscmpgeinngqidvpqillf--gatistsctlykklfgsts-sfccll	206	
QY	114 DG----WTNDIPICEVVKCLPYTABENGRKIVSSAMEPDRE-YHEGGAVRFCNSGYKIEG	168	
Db	207 sgsgswsdplpecreiyec-----papqdngilqgerdhrygrgsvtvacnkkgftlmig	261	
OY	169 DEEMHC---SDSGFMSKEPKC-----VEISCKSPDVNI-----GSPISOK	206	
Db	262 ehsjctcvmddegewsgppcecrkglslskvptlvqkpltnvpptevspsdq	315	
RESULT	8		
ID	W26317		
XX	W26317 standard; Protein: 381 AA.		
AC	W26317;		
XX	17-NOV-1997 (first entry)		
DT			
XX	Human decay accelerating factor.		
DE			
XX	DAF: decay accelerating factor: human: C3 complement inhibitor:		
KM	retrovirus; vector: gene therapy; stem cell.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	Peptide	1..34	
FT	/label= Sig_peptide		
FT	Protein	35..381	
FT	/label= Mat_protein		
XX			
PN	US5643770-A.		
XX			
PD	01-JUL-1997.		
XX			
PF	21-JUL-1994; 94US-0278630.		
XX			
PR	21-JUL-1994; 94US-0278630.		
XX			
PA	(ALEX-) ALEXION PHARM INC.		
PI	Mason JM, Squinto SP;		
XX			
DR	WPI: 1997-350243/32.		
N	N-PsDB: T84471.		
PT	Retroviral vector particle expressing complement inhibitor activity		
PT	- for transducing cells in body fluids containing complement		
XX			
PS	Disclosure: Column 41-46; 32pp; English.		
XX			
CC	This protein sequence comprises human decay accelerating factor		
CC	(DAF), a protein that has C3 complement inhibitor activity.		
CC	Claimed retroviral vector particles express a complement inhibitor		
CC	such as DAF, and are thereby protected from inactivation upon		
CC	exposure to body fluids containing complement. Also claimed are: (1)		
CC	a producer cell producing the vector particle; and (2) a chimeric		
CC	retroviral envelope protein (see also W26326) with at least part of		
CC	the N-terminal receptor-binding domain removed and replaced by a		
CC	protein domain having a complement inhibitor activity. The vector		
CC	is used in a claimed method for transducing cells in the presence of		
CC	a body fluid containing complement, preferably ex vivo, especially		
CC	for gene therapy, e.g. of hereditary or acquired blood disorders by		

CC	transduction of haematopoietic stem cells.					
XX	Sequence	381 AA;				
S0						
	Query Match	22.4%; Score 261; DB 18; Length 381;				
	Best Local Similarity	29.9%; Pred. No. 1.4e-16;				
	Matches	70; Conservative	84; Mismatches	84; Indels	38; Gaps	12;
OY	1 EDCN--EIPPRNRTEILGWSMDQTY-PEGTQAIVKCRPGTV---SIGNVIMVCRKGW	53				
	:   :   :   :   :   :   :   :   :   :   :   :					
Db	92 efctnscveptlrlnasalkpyitqyifvgtvveyecrpsyrrepslktclqnklw	151				
OY	54 VALNLPRKCQRKPCGHPGDTFRCTTLTGNGVFEEYGAVAYCNGGYOLLGEINRRECDT	113				
	:   :   :   :   :   :   :   :   :   :   :   :   :					
Db	152 --slavelfckkkkscnpgeirngqidvpqllf--gatlstiscnlgylfgsts-sfcil	206				
OY	114 DC----WTWDIPICEVVKCLPTAPENCKIVSSAMEPDRE-YHFGQAVRFVCNSGKIENG	168				
	:   :   :   :   :   :   :   :   :   :   :   :   :					
Db	207 sgssvsgwdplpeccreiy-----papqdldngligerdhygrqsvtyacnkgftmrg	261				
OY	169 DEEMHC--SDDGFWSEKPKC-----VEISCKSPDIVN----GSPISQK	206				
	: :   :   :   :   :   :   :   :   :   :   :   :					
Db	262 ensiyctvnndegewsgpppecrgxksltakvptvgkrtvtuvptlevsptsqk	315				
RESULT	9					
X31740	ID	Y31740 standard; Protein; 381 AA.				
XX						
AC	Y31740;					
XX						
DE	22-NOV-1999	(first entry)				
XX						
XX	Human CD55 and 791Tgp72 tumour associated antigen.					
KM	CD55; decay accelerating factor; DAF; tumour associated antigen;					
KW	791Tgp72; colorectal cancer; breast cancer; ovary cancer;					
XX	osteosarcoma; vaccine; human.					
OS	Homo sapiens.					
XX						
FH	Key	Location/Qualifiers				
FT	Peptide	1..34				
FT		/note= "signal peptide"				
FT	Protein	35..380				
XX		/note= "mature protein"				
PN	MO9943800-A1.					
XX						
PD	02-SEP-1999.					
PF	26-FEB-1999;	99MO-GB00582.				
XX						
PR	26-FEB-1998;	98GB-0004065.				
XX						
PA	(CANC-) CANCER RES CAMPAIGN TECHNOLOGY.					
XX						
PI	Durant LG, Spendlove I;					
XX						
DR	WPI: 1999-540585/45.					
XX	N-PsDB; X87914, X87915.					
PT	Cancer vaccine containing CD55 family polypeptide, to induce at					
PT	least one of T helper, cytotoxic T cell or natural killer immune					
XX	response					
PS	Claim 4; Fig 10; P 82pp; English.					
CC	The present sequence represents a new human tumour associated					
CC	antigen, 791Tgp72, which is over-expressed on a wide range of solid					
CC	tumours. The amino acid sequence of 791Tgp72 is identical to that					
CC	of human CD55, although the glycosylation patterns of the 2					



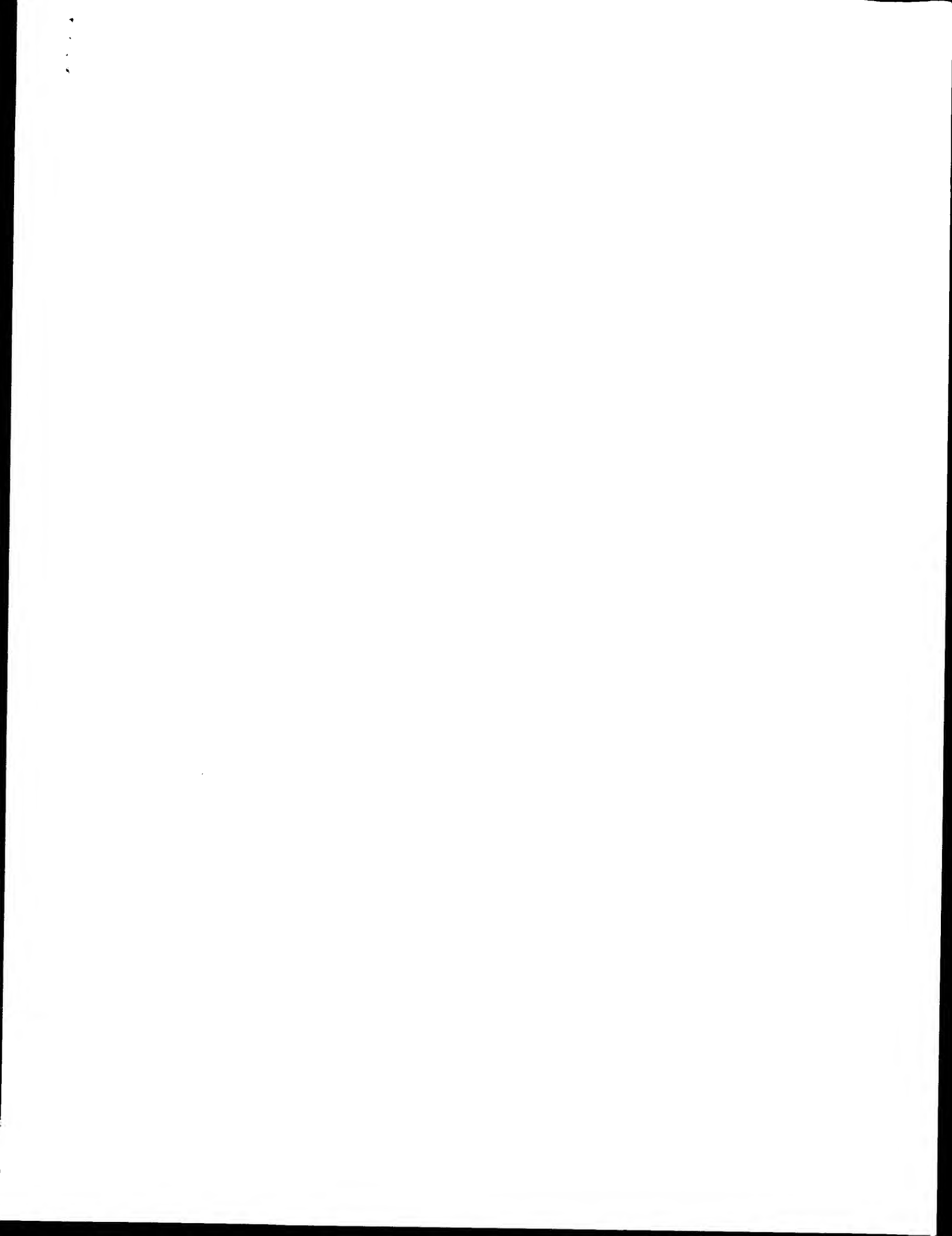




(e.g. kits for determining whether a patient has an abnormality

WPI; 1992-375009/46.





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:57 ; Search time 27.94 Seconds  
(without alignments)  
236.694 Million cell updates/sec

Title: US-09-316-163-9

Perfect score: 1163  
Sequence: 1 EDCNLEPPRRNTEILGWSMS.....VEISCKSPDVINGSPISQKI 207

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first=45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1163	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	822	70.7	1234	1 CFAH_MOUSE	P06909 mus musculu
3	261	22.4	381	1 DAF_HUMAN	P08174 homo sapien
4	255	21.9	2039	1 CRI_HUMAN	P17927 homo sapien
5	253.5	21.8	484	1 LEM2_PIG	P18110 sus scrofa
6	251	21.6	340	1 DAF_PONY	P49457 pongo pygma
7	246.5	21.2	507	1 DAF_CAVPO	Q60401 cavia porce
8	246	21.2	263	1 VCP_VACCV	P10998 vaccinia vi
9	244	21.0	597	1 CAPP_HUMAN	P04003 homo sapien
10	238.5	20.5	377	1 MCP_HUMAN	P15529 homo sapien
11	236	20.3	360	1 CAPH_HSVSA	Q01016 herpesvirus
12	231.5	19.9	390	1 DAF1_MOUSE	Q61475 mus musculu
13	230.5	19.8	610	1 LEM2_HUMAN	P16581 homo sapien
14	228	19.6	469	1 CAPP_MOUSE	P08607 mus musculu
15	227.5	19.6	788	1 LEM3_MOUSE	Q01102 mus musculu
16	227.5	19.6	830	1 LEM3_HUMAN	P16109 homo sapien
17	226	19.4	549	1 LEM2_RAT	P98105 rattus norv
18	222.5	19.1	407	1 DAF2_MOUSE	Q61476 mus musculu
19	222	19.1	768	1 LEM3_RAT	P98106 rattus norv
20	220	18.9	611	1 LEM2_CANFA	P33730 canis faml
21	218.5	18.8	612	1 LEM2_MOUSE	Q00690 mus musculu
22	216	18.6	558	1 CAPP_RAT	Q63514 rattus norv
23	214.5	18.4	551	1 LEM2_RABIT	P27113 oryctolagus
24	214.5	18.4	610	1 CAPP_BOVIN	Q28085 bos taurus
25	212	18.2	769	1 LEM3_SHEEP	Q28085 bos taurus
26	212	18.2	1025	1 CR2_MOUSE	P19070 ovis aries
27	210.5	18.1	646	1 LEM3_BOVIN	P42201 bos taurus
28	210.5	18.1	1033	1 CR2_HUMAN	P20023 homo sapien
29	210	18.1	372	1 LEM1_MOUSE	P18137 mus musculu
30	205.5	17.7	465	1 LEM2_BOVIN	P98107 bos taurus
31	202	17.4	958	1 APOH_BOVIN	P17650 bos taurus
32	202	17.4	958	1 HIG_DROME	Q09101 drosophila
33	200	17.2	372	1 LEM1_RAT	P30836 rattus norv

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	1231 AA.
1	CFAH_HUMAN			
ID	P08603			
AC	01-AUG-1988 (rel. 08, Created)			
DT	01-JAN-1990 (rel. 13, Last sequence update)			
DE	30-MAY-2000 (rel. 39, Last annotation update)			
OC	COMPLEMENT FACTOR H PRECURSOR.			
GN	HE1 OR HF OR CEH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER:			
RX	MEDLINE: 86134059.			
RA	Ripchoe J., Day A.J., Harris T.J.R., Sim R.B.;			
RT	"The complete amino acid sequence of human complement factor H.";			
RL	Biochem. J. 249:593-602(1988).			
RN	[2]			
RP	SEQUENCE OF 53-445 FROM N.A.			
RX	MEDLINE: 87054207.			
RA	Schulz T.F., Schwebble W., Stanley K.K., Weiss E., Dierich M.P.;			
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA			
RL	sequence of the 38-kDa tryptic fragment containing the binding site			
RL	for C3b.";			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
RN	[3]			
RP	SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE: 86169701.			
RA	Kristensen T., Wetsel R.A., Tack B.F.;			
RT	"Structural analysis of human complement protein H: homology with C4b			
RL	binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";			
RL	J. Immunol. 136:3407-3411(1986).			
RN	[4]			
RP	SEQUENCE OF 1047-1231 FROM N.A.			
RX	MEDLINE: 91201892.			
RA	Estaller C., Koistinen V., Schwebble W., Dierich M.P., Weiss E.H.;			
RT	"Cloning of the 1.4-kb mRNA species of human complement factor H			
RL	reveals a novel member of the short consensus repeat family related			
RL	to the carboxy terminal of the classical 150-kDa molecule.";			
RL	J. Immunol. 146:3190-3196(1991).			
RN	[5]			
RP	SEQUENCE OF 19-35.			
RX	MEDLINE: 83048213.			
RA	Sim R.B., Discipio R.G.;			
RT	"Purification and structural studies on the complement-system control			
RL	protein beta 1H (Factor H).";			
RL	Biochem. J. 205:285-293(1982).			
RN	[6]			
RP	STRUCTURE BY NMR OF 927-985 (SUSHI 16).			
RX	MEDLINE: 91278097.			
RA	Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;			
RT	"three-dimensional structure of a complement control protein module			
RL	in solution.";			
RL	J. Mol. Biol. 219:717-725(1991).			

Q01339 mus musculu  
P14151 homo sapien  
Q95198 macaca mula  
Q95237 pan troglod  
Q28768 papio hamad  
Q95235 pongo pygma  
P05160 homo sapien  
Q63515 rattus norv  
P28175 tachyples  
P02749 homo sapien  
Q26422 carlinoscor  
P33703 canis faml

```

RN [7]
RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).
RX MEDLINE; 92232649.
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,
RT "Discol P.C., Sim B., Campbell I.D.;"
RT "Solution structure of the fifth repeat of factor H: a second example
RL of the complement control protein module.;"
RL Biochemistry 31:3626-3634(1992).
RN [8]
RP STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).
RX MEDLINE; 93323119.
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,
RT "Sim B., Campbell I.D.;"
RT "Solution structure of a pair of complement modules by nuclear
magnetic resonance.;"
RL J. Mol. Biol. 232:268-284(1993).
CC -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3)NBB COMPLEX (C5
CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION
CC 341 ONWARD DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00716; CA68704.1; -
DR EMBL; X04697; CAB41739.1; ALT_FRAME.
DR EMBL; M65294; AAA35948.1; -
DR PIR; S00254; NBH0H.
DR PIR; S03013; S03013.
DR PIR; S00254; S00254.
DR PDB; IHCC; 15-APR-92.
DR PDB; IHPH; 15-JUL-93.
DR PDB; IHP1; 15-JUL-93.
DR MIM; 134370; -
DR INTERPRO: IPR000436; -
DR PFAM: PF00084; sushi. 20.
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;
KW Signal; 3d-structure; Polymorphism.
FT SIGNAL 1 18
FT CHAIN 19 1231
FT DOMAIN 20 1229
FT REPEAT 20 81
FT REPEAT 84 142
FT REPEAT 145 206
FT REPEAT 209 263
FT REPEAT 266 321
FT REPEAT 324 386
FT REPEAT 388 443
FT REPEAT 447 506
FT REPEAT 508 565
FT REPEAT 568 624
FT REPEAT 629 685
FT REPEAT 690 745
FT REPEAT 752 804
FT REPEAT 810 865
FT REPEAT 869 927
FT REPEAT 930 985
FT REPEAT 988 1044
FT REPEAT 1047 1103
FT REPEAT 1108 1164
FT REPEAT 1166 1229
FT DISULFID 21 66
FT DISULFID 52 80
FT DISULFID 85 129
FT DISULFID 114 141

```

```

FT DISULFID 146 192
FT DISULFID 178 205
FT DISULFID 210 251
FT DISULFID 237 262
FT DISULFID 267 309
FT DISULFID 294 320
FT DISULFID 325 374
FT DISULFID 357 385
FT DISULFID 389 431
FT DISULFID 416 442
FT DISULFID 448 494
FT DISULFID 477 505
FT DISULFID 509 553
FT DISULFID 536 564
FT DISULFID 597 623
FT DISULFID 597 611
FT DISULFID 630 673
FT DISULFID 659 684
FT DISULFID 691 733
FT DISULFID 719 744
FT DISULFID 753 792
FT DISULFID 781 803
FT DISULFID 811 853
FT DISULFID 839 864
FT DISULFID 870 915
FT DISULFID 901 926
FT DISULFID 931 973
FT DISULFID 959 984
FT DISULFID 989 1032
FT DISULFID 1018 1043
FT DISULFID 1048 1091
FT DISULFID 1077 1102
FT DISULFID 1109 1152
FT DISULFID 1138 1163
FT DISULFID 1167 1218
FT DISULFID 1201 1228
FT CARBOHYD 529 718
FT CARBOHYD 718 718
FT CARBOHYD 802 802
FT CARBOHYD 822 822
FT CARBOHYD 882 882
FT CARBOHYD 911 911
FT CARBOHYD 1029 1029
FT CARBOHYD 1095 1095
FT VARIANT 402 402
FT CONFLICT 21 21
FT CONFLICT 30 30
FT CONFLICT 34 34
FT CONFLICT 53 54
FT STRAND 870 871
FT STRAND 876 876
FT TURN 877 878
FT STRAND 881 881
FT STRAND 890 891
FT TURN 893 894
FT STRAND 896 900
FT TURN 903 904
FT STRAND 907 907
FT STRAND 912 916
FT TURN 917 918
FT STRAND 919 920
FT STRAND 926 926
SQ SEQUENCE 1231 AA; 139125 MW; C65EC8CF800B3FD CRC64;

```

```

/FTID=VAR_001979.
C -> Q (IN REF. 3).
T -> V (IN REF. 3).
T -> Q (IN REF. 3).
RP -> IL (IN REF. 2).
H -> Y.

```

```

Query Match 100.0%; Score 1163; DB 1; Length 1231;
Best Local Similarity 100.0%; Pred. No. 1,2e-94;
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 EDCNLEPPRRNTEITIGSWSDQTYPEGQAIYKCRPGYSLGNVTMCRKGFWALNPLR 60
DB 19 EDCNLEPPRRNTEITIGSWSDQTYPEGQAIYKCRPGYSLGNVTMCRKGFWALNPLR 78

```









Query Match	Best local Similarity	Score 255;	DB 1;	Length 2039;
Matches 67;	Conservative 39;	Mismatches 83;	Indels 32;	Gaps 13;
3 CNE---	LPRRNTELTLSWSDDQYRPECTGATYKRCRPGYSLGNVIMWCCKKEWALNPL	59		
43 CNAEPMLEPAPRTNUT---	DEFEPPIDYIYNECEPBGSGRPSFIITLCKNSWGTGAND-	97		
60 RKCCRRPGGHDPEFGFTLTGNGVNEFYGYKAVYATCEGYOLLEINVRREC---	DRDG	115		
98 -RCRRKSCRRNDPPDPNGWVHIK--	IFGSGQIKYSCSKYGRLLIGS -SSATCIISGDYVI	153		
116 WTNDIPICEVAK-LPVTAPENGKIVSSAMDEDRYHFGQAVRFVCGNSG----	YKIEG	168		

```

Db      154 WDNETPICDIRPGLPLPTLT-TNCFISTNNR---NFHYGSVWYTRCPNGSGGKRVFELYG    209
Oy      169 DEEMHC--SDD--GFWSKEPKC-VELSCSPDYINSPTS    204
       :::|::|| |::|::|::|::|::|::|::|::|
Db      210 EPSTCTSNDOVGIVSWGPAQCIIIPKKTCPPNVENGIILVS    250


RESULT   5
LEM2_PIG STANDARD; PRT;         484 AA.
AC        LEM2_PIG
AP          P98I10;
DT     01-FEB-1996 (Rel. 33, Created)
DM     01-FEB-1996 (Rel. 33, Last sequence update)
DR     15-JUL-1998 (Rel. 36, last annotation update)
DE E-selectin precursor (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1)
DE (ELAM-1) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 2) (LECAM2)
DE (CD62E).
GN SELF.
OS sus scrofa (pig).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTIC ENDOTHELIUM;
RX MEDLINE; 9507139Z.
RA Rollins S.A., Evans M.J., Johnson K.K., Elliot E.A., Squinto S.P.,
RT Mats L.A., Rother R.P.;
RT "Molecular and functional analysis of porcine E-selectin reveals a
RL potential role in xenograft rejection.";
RM Biochem. Biophys. Res. Commun. 204:763-771(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-AORTIC ENDOTHELIUM;
RX MEDLINE; 94271236.
RA Tsang Y.T.M., Haskard D.O., Robinson M.K.;
RT "Cloning and expression kinetics of porcine vascular cell adhesion
molecule.";
RL Blochem. Biophys. Res. Commun. 201:805-805f(1994).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIMILY-Lewis x (ALPHA1->3)fucosylated DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN ACUTE CELLULAR ALLOGRAFT
CC REJECTION AND PROBABLY ALSO IN XENOGRFT REJECTION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: TO OTHER SELECTINS/LecamS.
CC -1- SIMILARITY: CONTAINS A C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-like domain.
CC -1- SIMILARITY: CONTRAINS 4 SUSHI (SCR) REPEATS; PORCINE E-LECTIN LACKS
CC THE HUMAN SUSHI-1 AND -4 EQUIVALENENTS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way com-
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensels@isb-sib.ch).
-----
DB      EMBL; L39076; AAA61545.1; -
DB      EMBL; U08350; AAA21541.1; -
DB      HSSP; PI6581; LESL.
DR INTERPRO; IPRO00436; -
DR INTERPRO; IPR000561; -
DR INTERPRO; IPR001304; -
DR INTERPRO; IPR002396; -
DR PFAM; PF00008; EGF_1.
DR PFAM; PF00059; lectin_c_1.
DR PFAM; PF00084; sushi_4.
DR PRINTS; PR00343; SELECTIN.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.

```

	PROSITE; PS00615; C-TYPE LECTIN.1; 1.
DR	PROSITE; PS00615; C-TYPE LECTIN.2; 1.
KM	Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW	Selectin; Signal; Sushi; Repeat.
FT	SIGNAL 1 22 POTENTIAL.
FT	CHAIN 23 484 E-SELECTIN.
FT	DOMAIN 23 429 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 430 451 POTENTIAL.
FT	DOMAIN 452 484 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 39 139 C-TYPE LECTIN (SHORT FORM).
FT	DOMAIN 140 176 EGF-LIKE.
FT	DOMAIN 180 421 4 X SUSHI (SCR) REPEATS.
FT	REPEAT 180 236 SUSHI 1.
FT	REPEAT 239 299 SUSHI 2.
FT	REPEAT 302 362 SUSHI 3.
FT	REPEAT 365 421 SUSHI 4.
FT	DISULFID 41 139 BY SIMILARITY.
FT	DISULFID 112 131 BY SIMILARITY.
FT	DISULFID 144 155 BY SIMILARITY.
FT	DISULFID 149 164 BY SIMILARITY.
FT	DISULFID 166 175 BY SIMILARITY.
FT	DISULFID 181 222 BY SIMILARITY.
FT	DISULFID 208 235 BY SIMILARITY.
FT	DISULFID 240 285 BY SIMILARITY.
FT	DISULFID 271 298 BY SIMILARITY.
FT	DISULFID 303 348 BY SIMILARITY.
FT	DISULFID 334 361 BY SIMILARITY.
FT	DISULFID 366 407 BY SIMILARITY.
FT	DISULFID 393 420 BY SIMILARITY.
FT	CARBOHYD 61 65 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 65 79 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 79 79 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 160 160 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 201 201 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 254 254 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 376 376 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CARBOHYD 400 400 N-LINKED (GLCNAG. . . ) (POTENTIAL).
FT	CONFLICT 253 253 C -> Y (IN REF. 2).
FT	CONFLICT 313 313 L -> F (IN REF. 2).
FT	CONFLICT 321 321 T -> N (IN REF. 2).
FT	CONFLICT 327 327 K -> N (IN REF. 2).
FT	CONFLICT 363 363 V -> A (IN REF. 2).
FT	CONFLICT 384 384 V -> M (IN REF. 2).
FT	CONFLICT 461 484 KPVSSESSCLOPGNSGYMPSDLI -> NIPLDAAPPAAPMDPTKCLLT (IN REF. 2).
FT	SEQUENCE 484 AA; 52567 MW; AFF74FE25CIFDD013 CRC64;
QY	Query Match 21.8%; Score 253.5; DB 1; Length 484;
DB	Best Local Similarity 28.6%; Pred. No. 3.8e-15;
MATCHES	56; Conservative 36; Mismatches 83; Indels 21; Gaps 9;
QY	22 QTYPECTOALYYKCRPGYSRLGNVIMYC-RKGEWVALNPLRKCKRKC---GHGDTPEGT 77
DB	196 OSIPWMNTTCFACEKEGEFLGPHEILOCTSGSSWDGRKP--TCAVVCIDTVGHFN--GD 250
QY	78 FTLLGVAVPEFYGXRAV--YVNCGEYOLLGEINVRREDTGS-WTNDPICEVVAYCLPVTPAP 134
DB	251 VSCHNSSIGTFAYKSTCFHFCABGFILOGAQL-ECTAQOGMTQOAPVCAKVAWCRAVSPQR 309
QY	135 ENGIIVSSAAEPREYHFQGAVPVNCNGYIKBGDEMHCSDDGFSKSKPKCVELISCKS 194
DB	310 KNG-LYVFHTSPITGEFLFYKSSCAFSCEBGEFLGSAQLACTSOGOMTQEVPSCOVVOCCSS 368
QY	195 PDV-----INGSPI 203
DB	369 LEVPREINMSCSGEPV 384
RESULT	6
DAF_PONPY	STANDARD;
ID_DAF_PONPY	PRT; 340 AA.
AC	P49457;



DR	EMBL	D55661	BAA09514.1	JOINED.	
DR	EMBL	D55662	BAA09514.1	JOINED.	
DR	EMBL	D55663	BAA09514.1	JOINED.	
DR	EMBL	D55664	BAA09514.1	JOINED.	
DR	EMBL	D55667	BAA09515.1	-	
DR	EMBL	D55656	BAA09515.1	JOINED.	
DR	EMBL	D55657	BAA09515.1	JOINED.	
DR	EMBL	D55658	BAA09515.1	JOINED.	
DR	EMBL	D55659	BAA09515.1	JOINED.	
DR	EMBL	D55660	BAA09516.1	JOINED.	
DR	EMBL	D55661	BAA09516.1	JOINED.	
DR	EMBL	D55662	BAA09516.1	JOINED.	
DR	EMBL	D55663	BAA09516.1	JOINED.	
DR	EMBL	D55664	BAA09516.1	JOINED.	
DR	EMBL	D55665	BAA09516.1	JOINED.	
DR	EMBL	D55667	BAA09517.1	-	
DR	EMBL	D55656	BAA09517.1	JOINED.	
DR	EMBL	D55657	BAA09517.1	JOINED.	
DR	EMBL	D55658	BAA09517.1	JOINED.	
DR	EMBL	D55659	BAA09517.1	JOINED.	
DR	EMBL	D55660	BAA09517.1	JOINED.	
DR	EMBL	D55662	BAA09517.1	JOINED.	
DR	EMBL	D55663	BAA09517.1	JOINED.	
DR	EMBL	D55664	BAA09517.1	JOINED.	
DR	EMBL	D55665	BAA09517.1	JOINED.	
DR	EMBL	D55666	BAA09518.1	-	
DR	EMBL	D55667	BAA09518.1	JOINED.	
DR	EMBL	D55656	BAA09518.1	JOINED.	
DR	EMBL	D55657	BAA09518.1	JOINED.	
DR	EMBL	D55658	BAA09518.1	JOINED.	
DR	EMBL	D55659	BAA09518.1	JOINED.	
DR	EMBL	D55660	BAA09518.1	JOINED.	
DR	EMBL	D55661	BAA09518.1	JOINED.	
DR	EMBL	D55662	BAA09518.1	JOINED.	
DR	EMBL	D55663	BAA09518.1	JOINED.	
DR	EMBL	D55664	BAA09518.1	JOINED.	
DR	EMBL	D55665	BAA09518.1	JOINED.	
DR	EMBL	D55666	BAA09518.1	JOINED.	
DR	EMBL	D49416	BAA08397.1	-	
DR	EMBL	D49417	BAA08397.1	-	
DR	EMBL	D49418	BAA08398.1	-	
DR	EMBL	D49419	BAA08399.1	-	
DR	EMBL	D49420	BAA08400.1	-	
DR	EMBL	D49421	BAA08400.1	-	
DR	HSSP	P08603	1HCC		
DR	INTERPRO	IPR000436	-		
DR	PFAM	PF000084	sush1.4		
KW	Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor; Alternate splicing; Signal; Sush1.				
FT	SIGNAL	1	31	POTENTIAL.	
FT	CHAIN	32	451	COMPLEMENT.	
FT	PROPEP	452	507	DECAY-ACCELERATING FACTOR.	
FT	REPEAT	32	283	COMPLEMENT IN MATURE FORM (BY SIMILARITY)	
FT	REPEAT	32	93	4 X SUSH1 (SCR) REPEATS.	
FT	REPEAT	94	158	SUSH1 1.	
FT	REPEAT	159	220	SUSH1 2.	
FT	REPEAT	221	283	SUSH1 3.	
FT	DOMAIN	284	458	SUSH1 4.	
FT	LIPID	451	451	SER/THR-RICH.	
FT	CARBOHYD	57	57	GPI-ANCHOR (BY SIMILARITY).	
FT	CARBOHYD	93	93	N-LINKED (GLCNAC. . .) (POTENTIAL).	

[illegible]



```

RP SEQUENCE OF 49--88.
RX MEDLINE: 85296001.
RA Chung L.P., Gagnon J., Reid K.B.M.:
RT "Amino acid sequence studies of human C4b-binding protein: N-terminal
RT sequence analysis and alignment of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment."
RL Mol. Immunol. 22:427-435(1985).
RN 171
RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.
RA MEDLINE: 83221615.
RX Dahlbeck B., Smith C.A., Mueller-Eberhard H.J.:
RT "Visualisation of human C4b-binding protein and its complexes with
RT vitamin K-dependent protein S and complement protein C4b."
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COPACOR TO C3b/C4b INACTIVATOR
CC (C3bINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4b. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4bC2a COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2a. ALPHA
CC CHAIN BINDS C4b. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 KDA HOMOPOLYMER OF ALPHA CHAINS OR A 500 KDA
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4b AT THE END.
CC -1- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isp-slb.ch/annouce/
CC or send an email to license@isp-slb.ch).
CC -----
DR EMBL: M31452; AAA36507.1; -.
DR EMBL: M62486; AAA36506.1; -.
DR EMBL: M62475; AAA36506.1; JOINED.
DR EMBL: M62476; AAA36506.1; JOINED.
DR EMBL: M62477; AAA36506.1; JOINED.
DR EMBL: M62478; AAA36506.1; JOINED.
DR EMBL: M62479; AAA36506.1; JOINED.
DR EMBL: M62480; AAA36506.1; JOINED.
DR EMBL: M62481; AAA36506.1; JOINED.
DR EMBL: M62482; AAA36506.1; JOINED.
DR EMBL: M62484; AAA36506.1; JOINED.
DR EMBL: M62485; AAA36506.1; JOINED.
DR EMBL: X04284; CAB51244.1; -.
DR EMBL: X04296; CAA27839.1; -.
DR EMBL: X02865; CAA26617.1; -.
DR PIR: A33568; NBHUC4.
DR HSSP: P10998; LVVC.
DR MIM: 120830; -.
DR INTERPRO: IPR000436; -.
DR PFAM: PF000084; sush1; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal;
KW polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597 C4b-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 1.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.

```

Query Match	Best Local Similarity	Score	DB 1:	Length	597:
Matches	59; Conservative	29.5%;	Pred. No. 3.3e-14;	Mismatches 77;	Indels 30; Gaps 10;
20	SPQTEGQATYKCRPGY-RSLGNYIYVCRK-GEWVALNPLKCRKRCGHPGDPFCT	77	21.0%;	DB 1:	Length 597;
67	TETREFTGTTLTKYTCLPAGYVRSHSTGTLCNSDGEWV-YNTF--CIYKCRHPGE-----	118	29.5%;	Pred. No. 3.3e-14;	Mismatches 77;
78	FTLTGNNV-----FEYGVAAVYTCNGCYQLLGEINRECDTD-----GWTNDPICEVWC	128	21.0%;	DB 1:	Length 597;
119	--LRNQVEVETKIDLSFGSIOIEFSGSGFPLIGSTTSR-CEVDQGVGWSHPLEPQCEIVKC	175	29.5%;	Pred. No. 3.3e-14;	Mismatches 77;
129	LPTAENCKIYSSAMEPREYHFGQAVFVNCSGKIKEDDEMHSCSD---GFWSKEK	184	21.0%;	DB 1:	Length 597;
176	KPPDIRNGHSE-----ENFYVIGESVYISCDPRSLGHASISCTVENETIGVWRPSP	231	29.5%;	Pred. No. 3.3e-14;	Mismatches 77;
185	PKCVEISCKSPDYINGSPIS	204	21.0%;	DB 1:	Length 597;
232	PCEKITCRKPDVSHGEMVS	251	29.5%;	Pred. No. 3.3e-14;	Mismatches 77;
RESULT 10					
MCP_HUMAN					
AC	P15529;				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	MEMBRANE COPACITOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST				
DE	LEUCOCYTE COMMON ANTIGEN) (TLX).				
GN	MCP.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
RN	(1)				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.				
RX	MEDLINE; 88286080.				
RA	Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,				
RA	Rebittsch M.B., Lemons R.S., Seya T., Atkinson J.P.;				
RT	"Molecular cloning and chromosomal localization of human membrane				
RT	cocfactor protein (MCP). Evidence for inclusion in the multigene				
RT	family of complement-regulatory proteins."				



RL J. Exp. Med. 168:181-194(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RX MEDLINE; 93119658.  
 RA Cervoni F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.;  
 RT "Characterization of a cDNA clone coding for human testis membrane  
 RT cofactor protein (MCP, CD46).";  
 RL Mol. Reprod. Dev. 34:107-113(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RX MEDLINE; 91267562.  
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,  
 RA McKenzie I.F.;  
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a  
 RT regulator of complement activation.";  
 RL Immunogenetics 33:335-344(1991).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RX MEDLINE; 94014356.  
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,  
 RA Kumar V.;  
 RT "Characterization of the promoter region of the membrane cofactor  
 RT protein (CD46) gene of the human complement system and comparison to  
 RT a membrane cofactor protein-like genetic element.";  
 RL J. Immunol. 151:4137-4146(1993).  
 RN [5]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE; 92289809.  
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;  
 RT "Tissue-specific and allelic expression of the complement regulator  
 RT CD46 is controlled by alternative splicing.";  
 RL Eur. J. Immunol. 22:1513-1518(1992).  
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST  
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY  
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD  
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3  
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH  
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT  
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE  
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE  
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN  
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE  
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND  
 CC TO THE TRANSCRIPTS C AND D. AND THE TRANSCRIPTS E AND F  
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME  
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I  
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS  
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICED EXON 13.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT  
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.  
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN  
 CC (PROBABLE).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD46 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/PROW/CD/CD46.HTM".  
 CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC EMBL; Y00651; CAA68675.1; -  
 CC EMBL; S51940; AAB24802.1; -  
 CC EMBL; M58050; AAA62833.1; -  
 CC EMBL; A18585; CAA01400.1; -  
 CC EMBL; S65879; AAD13968.1; -

DR PIR; S01896; S01896.  
 DR HSSP; P10998; 1VVC.  
 DR MIM; 120920; -  
 DR INTERPRO; IPR000436; -  
 DR PFAM; PF00084; sush1; 4.  
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;  
 KW sush1; Alternative splicing.  
 FT SIGNAL 1 34  
 FT CHAIN 35 377  
 FT DOMAIN 35 328  
 FT TRANSMEM 329 351  
 FT DOMAIN 352 377  
 FT DOMAIN 35 284  
 FT REPEAT 35 95  
 FT REPEAT 98 158  
 FT REPEAT 161 224  
 FT REPEAT 227 284  
 FT DOMAIN 287 311  
 FT DISULFID 35 80  
 FT DISULFID 64 94  
 FT DISULFID 99 141  
 FT DISULFID 127 157  
 FT DISULFID 162 210  
 FT DISULFID 191 223  
 FT DISULFID 228 270  
 FT DISULFID 256 283  
 FT CARBOHYD 83 83  
 FT CARBOHYD 114 114  
 FT CARBOHYD 273 273  
 FT VARSPIC 286 300  
 FT VARSPIC 340 352  
 FT VARSPIC 353 361  
 FT VARSPIC 362 377  
 FT VARSPIC 362 377  
 FT VARSPIC 340 361  
 FT VARSPIC 339 339  
 SQ SEQUENCE 377 AA; 42247 MW; 2CA6F61752570B57 CRC64;  
 Query Match 20.5%; Score 238.5; DB 1; Length 377;  
 Best Local Similarity 29.6%; Pred. No. 6.1e-14;  
 Matches 64; Conservative 32; Mismatches 99; Indels 21; Gaps 9;  
 QY 2 DCELEPRRNTTEILTGSWSQDTYPEGTQALYKCRPGY--RSIGNVIMVCRKGEWALNP 58  
 DB 33 DACEPPPTFAMELLIGK-PKPYEIEGERVDYKCKKGFIYPLATHHTICDRNHTWLPVSD 91  
 QY 59 LRCKORPGHPGDTPEGFTLLGNAVFEYGVAVTTCNGYOLLG-EIYRCDTIDG-- 115  
 DB 92 -DACYETCTYIND-PLNGOAVPANGTYERGOMHPLCNGYLLIEELIY--CEIKGSV 147  
 QY 116 -WTNDIPCEVVKCLPYVAPENGKIVSSAMPEBREHFEGOAVFVNSG-----YKIEG 168  
 DB 148 AIWSKGPICEKVLCTPPPRKIKNGKHTFEVE---VFELYDNTYSCDPAFGDPISLIG 204  
 QY 169 DEEMHSDGDFWSEKPKVEISCKSPDYVINGSPIS 204  
 DB 205 ESTTYGDNWSWSRAPECKVYCRPPVYENGKQIS 240  
 RESULT 11  
 CCPH\_HSVSA  
 ID CCPH\_HSVSA STANDARD; PRT; 360 AA.  
 AC 001016;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE COMPLEMENT CONTROL PROTEIN HOMOLOG (CCPH).  
 GN 4 OR CCPH.  
 OS Herpesvirus saimiri (strain 11).



CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Gammathepesvirinae; Rhadinovirus.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA MEDLINE: 92333688.  
 RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
 RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
 RA Honess R.W.,  
 RA "Primary structure of the herpesvirus saimiri genome.";  
 RA J. Virol. 66:5047-5058(1992).  
 RN [2]  
 RP SIMILARITY TO CCP.  
 RA MEDLINE: 92260674.  
 RA Albrecht J.-C., Fleckenstein B.,  
 RT "New member of the multigene family of complement control proteins in  
 RT herpesvirus saimiri.";  
 RL J. Virol. 66:3937-3940(1992).  
 CC -1- ALTERNATIVE PRODUCTS: A MEMBRANE-BOUND FORM AND A SECRETED FORM  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA).  
 CC -1- SIMILARITY: CONTAINS 3 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X64346; CAA45626.1; -;  
 DR EMBL: X64346; CAA45627.1; -;  
 DR EMBL: X60283; CAA42823.1; -;  
 DR EMBL: X60283; CAA42822.1; -;  
 DR PIR: B42534; WMBE2E.  
 DR PIR: A42534; WMBE1E.  
 DR PIR: S24567; S24567.  
 DR HSSP: P10998; IYVC.  
 DR INTERPRO: IPR000436; -;  
 DR PFAM: PF00084; sush1; 4.  
 DR Signal; Repeat; Sush1; Transmembrane; Alternative splicing;  
 KW Glycoprotein.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 360 COMPLEMENT CONTROL PROTEIN HOMOLOG.  
 FT DOMAIN 83 265 3 X SUSHI (SCR) REPEATS.  
 FT REPEAT 83 143 SUSHI 1.  
 FT REPEAT 146 206 SUSHI 2.  
 FT REPEAT 209 265 SUSHI 3.  
 FT TRANSMEM 328 350 POTENTIAL.  
 FT DISULFID 84 125 BY SIMILARITY.  
 FT DISULFID 111 142 BY SIMILARITY.  
 FT DISULFID 147 191 BY SIMILARITY.  
 FT DISULFID 175 205 BY SIMILARITY.  
 FT DISULFID 210 252 BY SIMILARITY.  
 FT DISULFID 238 264 BY SIMILARITY.  
 FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 289 302 R1CNGNCTTSMPTQ -> AECAPGSNTPISS (IN  
 FT SHORT ISOFORM).  
 FT VARSPLIC 303 360 MISSING (IN SHORT ISOFORM).  
 FT SEQUENCE 360 AA; 40006 MW; 6278A6C2ECD49669 CRC64;

DB 44 YNGTTLHTVTCDEGYAKRVPQVITVCYNGMWTV---PKKOKKKCSPPDLLNGRIYTV-G 99  
 QY 84 NVEFYGVKAVTCNNEGYYQLLEIN---YECDDGWTNDIPICEYVKCLPYAPENGKI 139  
 DB 100 NLY-VGSVITVTCNSGYSILGTSYACLLKRGGRVMTFRPPICDKKKPPQIANG-- 156  
 QY 140 VSSAMPEDEHYFGCAVRVCNSGK--IEGDEHMCSDGFW-SKEKPKVEISCKSPD 196  
 DB 157 --THTNVKDFYTYLDITVYSCNDETKLTLTGDSKICSEFGSVWNGETKCEFIKLPQ 214  
 QY 197 VIN 199  
 DB 215 VAN 217  
 RESULT 12  
 ID DAFL\_MOUSE STANDARD; PRT: 390 AA.  
 AC 061475: 061397; P97732;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE COMPLEMENT DECAY-ACCELERATING FACTOR, GPI-ANCHORED PRECURSOR  
 DE (DAF-GPI).  
 GN DAFL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;  
 RX MEDLINE: 95403982;  
 RA Spicer A.P., Seidlin M.F., Gendler S.J.;  
 RT "Molecular cloning and chromosomal localization of the mouse decay-  
 RT accelerating factor genes. Duplicated genes encode  
 RT glycosylphosphatidylinositol-anchored and transmembrane forms.";  
 RL J. Immunol. 155:3079-3091(1995).  
 RN [2]  
 RP SEQUENCE OF 7-390 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=SPLEEN;  
 RX MEDLINE: 96362213;  
 RA Fukuko Y., Yasui A., Okada N., Okada H.;  
 RT "Molecular cloning of murine decay accelerating factor by  
 RT immunoscreening.";  
 RT Int. Immunol. 8:379-385(1996).  
 RL -1- FUNCTION: PROTECTION OF CELLS FROM COMPLEMENT-MEDIATED DAMAGE (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.  
 CC -1- TISSUE SPECIFICITY: BRAIN, SECRETORY EPITHELIA, SKELETAL MUSCLE,  
 CC LIVER, TESTES, THYMUS, SPLEEN AND LYMPH NODE.  
 CC -1- DOMAIN: THE FIRST SUSHI DOMAIN (SCR1) IS NOT NECESSARY FOR THE  
 CC FUNCTION. SCR2 AND SCR4 PROVIDE THE PROPER CONFORMATION FOR THE  
 CC ACTIVE SITE ON SCR3 (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE RECEPTORS OF COMPLEMENT ACTIVATION  
 CC (RCA) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: I41366; AAB00091.1; -;  
 DR EMBL: D63679; BAA09830.1; -;  
 DR HSSP: P08603; 1HCC.  
 DR MCD; MG1:104850; DAFL.  
 DR INTERPRO: IPR000436; -;  
 DR PFAM: PF00084; sush1; 4.  
 KW Complement pathway; Glycoprotein; Membrane; Repeat; GPI-anchor;

Query Match 20.3%; Score 236; DB 1; Length 360;  
 Best Local Similarity 33.3%; Pred. No. 9, 6e-14;  
 Matches 61; Conservative 17; Mismatches 89; Indels 16; Gaps 7;  
 24 YPBGTAIVKRCRGRVSLGVIMVCRKGEWVALNPLKCKQKRCRGGHGPDPFGFTLTGG 83



```

FT CARBOHYD 179 179 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 199 199 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 203 203 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 265 265 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 312 312 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 332 332 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 503 503 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT CARBOHYD 527 527 N-LINKED (GLCNAC . .) (POTENTIAL) .
FT VARIANT 149 149 S -> R (ASSOCIATED WITH A RISK FACTOR FOR CAD).
FT FTID=VAR_004191.
FT CONFLICT 468 468 H -> Y (IN REF. 2).
FT SO SEQUENCE 610 AA: 66655 MW: 7043EE3C0D1229229 CRC64:
Query Match 19.8% Score 230.5 DB 1: Length 610:
Best Local Similarity 27.5% Pred. No. 5,2e-13:
Matches 53: Conservative 30: Mismatches 81: Indels 29: Gaps 10:
QY 23 TYPECTOALIKCRPGYSRLGNYIMVC-RKGEVVALNPLRK---CQ--KRPE-----CGHP 70
Db 261 SEPMMTTCFCFDEEGBELMGAQLSQTSSGMNDNEKPICKATCAVRAPRQNGSVRCSH- 319
QY 71 GDTPGTFTLITGGNVFEGYGVKAIVICNCSYOLGELINTRECDTG-WINDIPICEYVKCL 129
Db 320 --SPAGEFT-----FKSCNCFCEEGFEMLOGPAOV-ECTTGGQMTIQIIVECAFQCT 368
QY 130 PYTAENCKIVASMEPDREHFHGOAVRPVCSNGSKIECEDDEMHOSDSGFSEKEPKCYE 189
Db 369 ALSNERG-YNNCLPSASGSFRYGSGCPESCFOGYLVKSRILOCGPTGEMNEKPTECA 427
QY 190 ISCKSPDVINGSP 202
Db 428 VRC---DAVHOEP 437
RESULT 14
C4BP_MOUSE STANDARD: PRT: 469 AA.
AC P08607:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE CAB-BINDING PROTEIN PRECURSOR (C4BP).
GN C4BP OR C4BP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RX SEQUENCE FROM N.A. MEDLINE: 88024997.
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;
RT "cDNA structure of murine C4b-binding protein, a regulatory component
of the serum complement system.";
RL Biochemistry 26:4668-4674(1987).
CC -! FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
ACTIVATION. IT BINDS AS A CORACTOR TO C3B/C4B INACTIVATOR
(C3BIAN), WHICH THEN HYDRATIZES THE COMPLEMENT FRAGMENT C4B. IT
ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -! SUBUNIT: HOMODIMER; NOT COVALENTLY LINKED. MOUSE LACKS THE
BETA CHAIN OF C4BP.
CC -! SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -! SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.
CC -! CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```

or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17122; AAA37312.1; ALT_INIT.
DR PIR: A27117; NBMSCA.
DR HSSP: P10998; 1YVC.
DR MGD: MGI:88229; CAPP.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; SUSH1; 6.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sush1; Signal.
FT SIGNAL 1 56
FT CHAIN 57 469
FT DOMAIN 57 414 6 X SUSH1 (SCR) REPEATS.
FT REPEAT 57 116 SUSH1 1.
FT REPEAT 119 177 SUSH1 2.
FT REPEAT 180 241 SUSH1 3.
FT REPEAT 244 300 SUSH1 4.
FT REPEAT 302 356 SUSH1 5.
FT REPEAT 358 414 SUSH1 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA; 51551 MW; 41E137CB8D8C6321 CRC64;

Query Match 19.6%; Score 228; DB 1; Length 469;
Best Local Similarity 28.1%; Pred. No. 6.5e-13;
Matches 59; Conservative 40; Mismatches 87; Indels 24; Gaps 11;

QY 7 PPRRTTLLTSGWSMDT-YPECTQATYKCRGY-RSLGNVIMCR-KGEW-VALNPLKLC 62
DB 60 PPAIPNALPASDVNRCTFESHHTLTKYECPLGYGRISRMVYCKPSEMEWTSVS----C 115
QY 63 QKPCGHPGDPFGFTLTGNGVFEYGVKAVYTCNEGYYQLGLEINYREGDTDG---WTN 118
DB 116 AKKHCNRNGYLDNG---YVNGETITFGSOIEFSCQEGFLVNS-STSCSEVNRKGVAMSN 171
QY 119 DIPICEVVKCLPVTAPENGKIVSSAMEPDRREYHGOAVRFVCSNGYKTRGDEMHCS--- 175
DB 172 PEPCEVIYKCGPPDISNGK--HSGTEDEFPYRNHG--ISTYCDPGRFVSGSPFICTVYN 227
QY 176 -DDGFWMSKEKPCVBISSCKSPVINGSPIS 204
DB 228 KTVPYWSSSPTECKIKCSQPIHLHGIVS 257

RESULT 15
LEMS_MOUSE STANDARD: PRT: 768 AA.
AC 001102;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE P-SELECTIN PROTECTOR (GRANULE MEMBRANE PROTEIN 140) (GMP-140) (PADGEM)
DE (CD62E) (LEUKOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 3) (LECAM3).
GN SELP OR GMP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC -----
RN Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RP [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 92340571.
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectin. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92345617.
RA Sanders W.E., Jr., Wilson R.W., Ballantyne C.M., Beaudet A.L.;
RT "Molecular cloning and analysis of in vivo expression of murine P-
RT selectin.";
RL Blood 80:795-800(1992).
CC -1- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUKOCYTES. THE LIGAND RECOGNIZED IS STALYIN-LEWIS X.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC THE CELL SURFACE.
CC -1- INDUCTION: BY TUMOR NECROSIS FACTOR ALPHA.
CC -1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 SUSH1 (SCR) REPEATS; MOUSE P-LECTIN LACKS
CC THE HUMAN SUSH1-2 EQUIVALENT.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87861; AAA40008.1; -.
DR EMBL: M72332; AAA37712.1; -.
DR PIR: A42755; A42755.
DR HSSP: P16109; 1FSB.
DR MGD: MGI:98280; SELP.
DR INTERPRO: IPR000436; -.
DR INTERPRO: IPR000561; -.
DR INTERPRO: IPR001304; -.
DR INTERPRO: IPR002396; -.
DR PFAM: PF00008; EGF_1.
DR PFAM: PF00059; lectin_c_1.
DR PFAM: PF00084; Sush1; 8.
DR PFAM: PF00084; Sush1; 6.
DR PRINTS: PR00343; SELECTIN.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_2; 1.
DR PROSITE: PS00615; C-TYPE_LECTIN_3; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sush1; Repeat; Lipoprotein; Palmitate.
FT SIGNAL 1 41
FT CHAIN 42 768
FT DOMAIN 42 709 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 710 733 POTENTIAL.
FT DOMAIN 734 768 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 158 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 159 195 EGF-LIKE.
FT DOMAIN 199 700 8 X SUSH1 (SCR) REPEATS.
FT REPEAT 199 258 SUSH1 1.
FT REPEAT 261 320 SUSH1 2.
FT REPEAT 323 382 SUSH1 3.
FT REPEAT 385 444 SUSH1 4.
FT REPEAT 447 506 SUSH1 5.
FT REPEAT 509 568 SUSH1 6.
FT REPEAT 579 638 SUSH1 7.

```

```

FT REPEAT 641 700 SUSHI 8.
FT DISULFID 60 158 BY SIMILARITY.
FT DISULFID 131 150 BY SIMILARITY.
FT DISULFID 163 174 BY SIMILARITY.
FT DISULFID 168 183 BY SIMILARITY.
FT DISULFID 185 194 BY SIMILARITY.
FT DISULFID 200 244 BY SIMILARITY.
FT DISULFID 230 257 BY SIMILARITY.
FT DISULFID 262 306 BY SIMILARITY.
FT DISULFID 292 319 BY SIMILARITY.
FT DISULFID 324 368 BY SIMILARITY.
FT DISULFID 354 381 BY SIMILARITY.
FT DISULFID 386 430 BY SIMILARITY.
FT DISULFID 416 443 BY SIMILARITY.
FT DISULFID 448 492 BY SIMILARITY.
FT DISULFID 478 505 BY SIMILARITY.
FT DISULFID 510 554 BY SIMILARITY.
FT DISULFID 540 567 BY SIMILARITY.
FT DISULFID 580 624 BY SIMILARITY.
FT DISULFID 610 637 BY SIMILARITY.
FT DISULFID 642 686 BY SIMILARITY.
FT DISULFID 672 699 BY SIMILARITY.
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 654 654 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 679 679 PALMITATE (BY SIMILARITY).
FT LIPID 745 745 ENDOCYTOSIS SIGNAL (PROBABLE).
FT SITE 756 759 A -> E (IN REF. 2).
FT CONFLICT 724 724
SQ SEQUENCE 768 AA; 83098 MW; E5173074D2F66E68 CRC64;

```

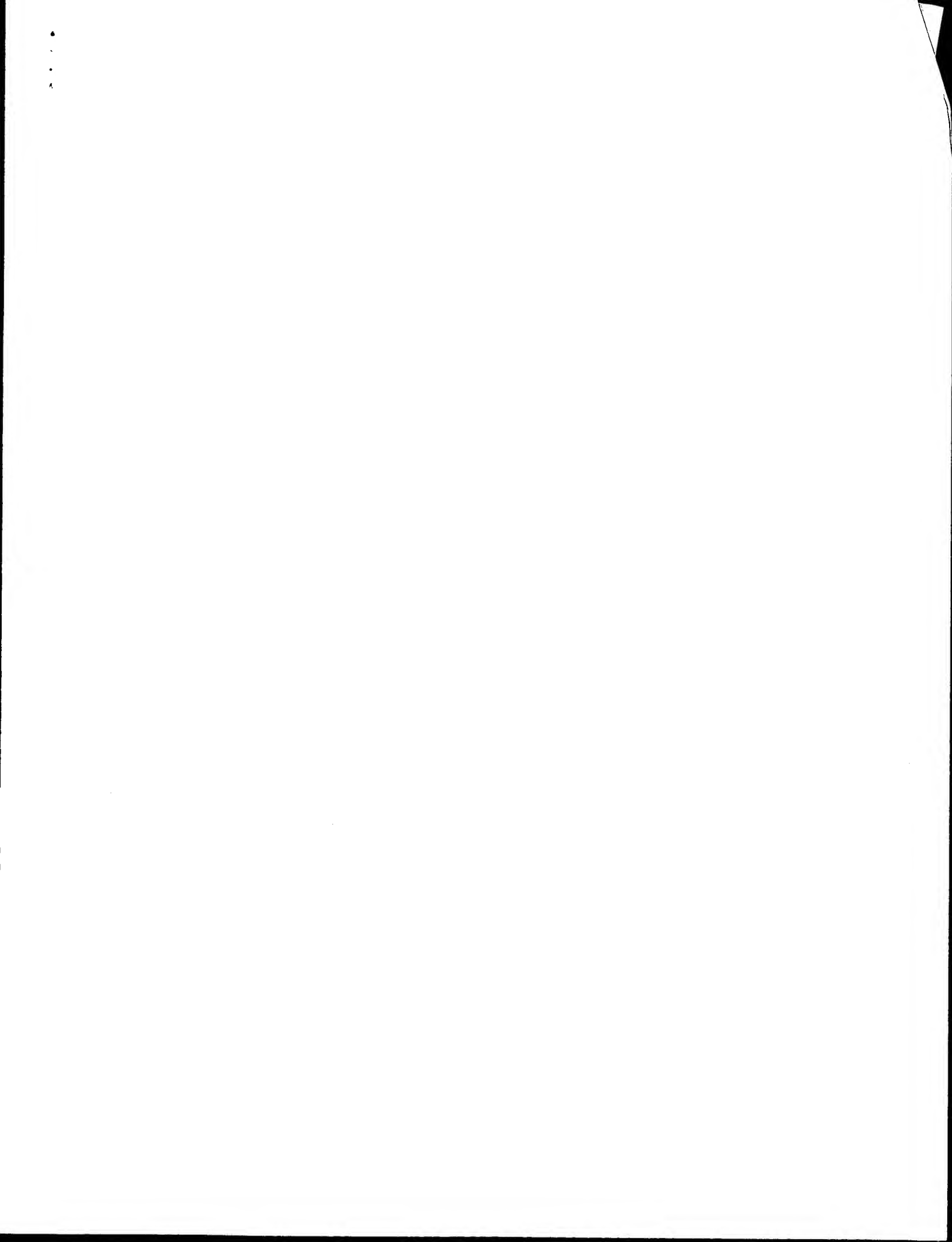
```

Query Match 19.6%; Score 227.5; DB 1; Length 768;
Best Local Similarity 26.4%; Pred. No. 1.2e-12;
Matches 61; Conservative 41; Mismatches 74; Indels 55; Gaps 13;

OY 1 EDCNELPRRNTTEITGMSDQ-----TYPEGTQAIYKCRPG--YRSLGNVIMCR 49
Db 129 QDCEVEIYIKSNS--APGKWNDEPCFKRRRALCYTASCQDMSCSNOGECIETIGSYTCSCY 186
OY 50 KG-----EWALNPLRRCOK-----RPGGHP-GDTPFGTFLTGNNVEYGVKAVY 94
Db 187 PGFYGPECEYV-----KECGKYNIPQHVLMNCSHPLGE-----FSFNSQCTF 228
OY 95 TCNEGYQL-LGEINVRCDTIDG-WTNDIPICEVVKCLPVTAPENGKIYSSAMEPDRHYH 151
Db 229 SCAREGYELDGPGL---QCLASGIWTNNPKCDAVQCOSLEAPPHGTM--ACMHPIAAFA 283
OY 152 FGQAVRFYCNAGSYKTEGDEBHCSDGDFWFSKPKCEIISCKSPDY-INGS 201
Db 284 YDSSCKFECCPGYRARGSNITLHCTGSGQWSEPLPTCEAIACEPPEIPIHGS 334

```

Search completed: November 21, 2000, 16:49:19  
 Job time: 322 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:22 ; Search time 70.54 Seconds  
(without alignments)  
274.007 Million cell updates/sec

Title: US-09-316-163-9  
Perfect score: 1163  
Sequence: 1 EDCNCLPPRRNTTELLGWSMS.....VEISCKSPDVINGSPISQKI 207

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP-SPREMBL-11:\*\*\*  
2: SP-SPREMBL-11:\*\*\*  
3: SP-fungi:\*\*\*  
4: SP-human:\*\*\*  
5: SP-invertebrate:\*\*\*  
6: SP-mammal:\*\*\*  
7: SP-mmc:\*\*\*  
8: SP-organella:\*\*\*  
9: SP-plant:\*\*\*  
10: SP-rodent:\*\*\*  
11: SP-virus:\*\*\*  
12: SP-vertebrate:\*\*\*  
13: SP-vertebrate:\*\*\*  
14: SP-unclassified:\*\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1163	100.0	449	4	Q14570
2	499	42.9	669	4	Q28085
3	368.5	31.7	1053	13	Q91275
4	276.5	23.8	645	12	Q9WRU2
5	275	23.6	645	12	Q9YTO8
6	270.5	23.3	550	12	P88903
7	270.5	23.3	550	12	Q40912
8	261	22.4	347	4	P78361
9	260	22.4	679	11	Q99254
10	256	22.0	522	6	Q28769
11	255.5	22.0	559	12	P87616
12	255	21.9	559	4	Q9UQV2
13	255	21.9	2039	4	Q16745
14	255	21.9	2489	4	Q16744
15	253.5	21.8	482	6	Q28982
16	248	21.3	315	6	Q28770
17	247	21.2	363	6	Q02839
18	247	21.2	661	6	Q29531
19	247	21.2	2014	6	Q29530

20	246.5	21.2	383	11	Q9WTR9	09wt19 cavia porce
21	243.5	20.9	497	11	Q63612	063612 rattus norv
22	243.5	20.9	559	11	Q63135	063135 rattus norv
23	243	20.9	222	6	Q19122	Q19122 callimico g
24	243	20.9	560	5	Q22328	Q22328 caenorhabdi
25	240.5	20.7	483	11	Q64735	Q64735 mus musculu
26	239.5	20.6	1911	6	Q29528	Q29528 papio hamad
27	238.5	20.5	349	4	Q15429	Q15429 homo sapien
28	238.5	20.5	378	6	Q62837	Q62837 saginus oe
29	237	20.4	222	6	Q19128	Q19128 plithicia pl
30	237	20.4	1652	5	Q9VIT9	Q9VIT9 dirosophila
31	236.5	20.3	1124	5	Q9VYR4	Q9VYR4 dirosophila
32	236	20.3	263	12	Q89061	Q89061 variola vir
33	236	20.3	263	12	Q89859	Q89859 variola vir
34	236	20.3	263	12	Q07033	Q07033 variola vir
35	235	20.2	263	12	Q89076	Q89076 variola vir
36	233.5	20.1	417	11	Q35520	Q35520 rattus norv
37	232	19.9	222	6	Q19125	Q19125 saimiri sci
38	232	19.9	314	6	Q62835	Q62835 saginus oe
39	232	19.9	377	6	Q62838	Q62838 saginus oe
40	231.5	19.9	285	6	Q19126	Q19126 macaca fasc
41	231.5	19.9	285	6	Q19127	Q19127 macaca mula
42	229.5	19.7	336	6	Q62834	Q62834 saginus oe
43	229.5	19.7	369	6	P79138	P79138 cercopithe
44	228.5	19.6	522	6	Q28797	Q28797 pan troglod
45	228	19.6	222	6	Q19124	Q19124 saginus my

#### ALIGNMENTS

RESULT 1  
ID Q14570 PRELIMINARY: PRT: 449 AA.  
AC Q14570: P88435:  
DT 01-NOV-1996 (TREMblrel. 01, created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88134059.  
RA Ripoche J., Day A.J., Harris T.J.R., Sim R.B.:  
RT "The complete amino acid sequence of human complement factor H";  
RL Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Tack B.F.:  
RT "Structural analysis of human complement protein H: homology with C4b  
binding protein, beta-2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RX VLK D.P., Williams S.A.:  
RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Dominguez O.:  
RT Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.  
RL EMBL: X07523; CAAS0403.1; -  
DR EMBL: M12383; AAAS2013.1; -  
DR EMBL: U56979; AAB01987.1; -  
DR EMBL: Z29665; CAA82763.1; -  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -  
DR PIRAM: PF00084; sushi; 7.  
KW SIGNAL.  
FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 449 POTENTIAL.  
SQ SEQUENCE 449 AA; 51033 MW; C2AAD47F155343E3 CRC64;

Query Match  
Best Local Similarity 100.0%; Score 1163; DB 4; Length 449;  
Matches 207; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCELPRTTELTLTSGMSDQTPGCAQYKCRPGYRSLGNVIMVCRKGEWALNPLR 60  
DB 19 EDCELPRTTELTLTSGMSDQTPGCAQYKCRPGYRSLGNVIMVCRKGEWALNPLR 78  
QY 61 KCQRPCGHPGDPFGFTLLTGNVFEYGVAVYTCNCGYOLLGEINRECDTDGWTNDI 120  
DB 79 KCQRPCGHPGDPFGFTLLTGNVFEYGVAVYTCNCGYOLLGEINRECDTDGWTNDI 138  
QY 121 PICEVYKCLPTAPENKIKYSAMEPDEHYHFGAVRVCNSGYKIEGDEMHCSDDGFW 180  
DB 139 PICEVYKCLPTAPENKIKYSAMEPDEHYHFGAVRVCNSGYKIEGDEMHCSDDGFW 198  
QY 181 SKKPKCVCISCKSPDVINGSPISOKI 207  
DB 199 SKKPKCVCISCKSPDVINGSPISOKI 225

## RESULT 2

Q28085 PRELIMINARY; PRT; 669 AA.

AC Q28085;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DE 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 96202005.  
RA Soames C.J., Day A.J., Sim R.B.;  
RT "Prediction from sequence comparisons of residues of factor H involved  
in the interaction with complement component C3b."  
RL Biochem. J. 315:523-531(1996).  
DR EMBL; X98697; CAA67257.1; -.  
DR HSSP; P10998; IYVC.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 11.  
FT NON\_TER 1  
FT NON\_TER 669  
SQ SEQUENCE 669 AA; 75683 MW; D0D9DB30EB747AC2 CRC64;

Query Match  
Best Local Similarity 42.9%; Score 499; DB 6; Length 669;  
Matches 85; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 76 GTFTLTGNVFEYGVAVYTCNCGYOLLGEINRECDTDGWTNDIPICEVYKCLPTAPE 135  
DB 3 GSPHLAEGNDFEGKAVYTCDEGQVMEAMFRCDTNGWTNDIPICEVYKCLPTAPE 62  
QY 136 NGKIYSSAMEPDEHYHFGAVRVCNSGYKIEGDEMHCSDDGFWSKKPKCVCISCKSP 195  
DB 63 NGKIYSSAMEPDEHYHFGAVRVCNSGYKIEGDEMHCSDDGFWSKKPKCVCISCKSP 122  
QY 196 DVINGSPISOK 206  
DB 123 VILNGQAVLPK 133

RESULT 3  
Q91275

ID Q91275 PRELIMINARY; PRT; 1053 AA.

AC Q91275;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT REGULATORY PLASMA PROTEIN.  
OS Paratubax nebulifer (barred sand bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;  
OC Perciformes; Percloides; Serranidae; Paratubrax.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE; 94318039.  
RA Dahmen A., Kaidoh T., Zipfel P.F., Giggli I.;  
RT "Cloning and characterization of a cDNA representing a putative  
nebulifer.";  
RT complement-regulatory plasma protein from barred sand bass (Paratubax  
nebulifer)."  
RL Biochem. J. 301:391-397(1994).  
DR EMBL; U21703; AAA92556.1; -.  
DR HSSP; P08603; IHEH.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 16.  
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match  
Best Local Similarity 31.7%; Score 368.5; DB 13; Length 1053;  
Matches 72; Conservative 25; Mismatches 76; Indels 9; Gaps 4;

QY 21 DQYRPESTQALYKCRPGYRSLGNVIMVCRKGEWALNPLRKQCRPGHPGDPFGFTL 80  
DB 47 EASTPGGRQVAVCGNAGYS--GFEKLVCEGKMETRG--AKCOPRSCGHPGDDQFAFPHL 102  
QY 81 TGGNVFEYGVAVYTCNCGYOLLGEINRECDTDGWTNDIPICEVYKCLPTAPENKIV 140  
DB 103 AEGNDFEYGSKVYTCQKQVMSRINRYRCVAGMGVGVPCVESQOCPLIHVNNOVI 162  
QY 141 SSAMEPDEHYHFGAVRVCNSGYKIEGDEMHCSDDGFWSKKPKCVCISCKSPDVIN 199  
DB 163 GG---PEEATFGNVAVFSCSKSEILDSPELYCDERGMSPVPCKAITCAPIEN 218  
QY 200 GS 201  
DB 219 GN 220

## RESULT 4

Q9WRU2 PRELIMINARY; PRT; 645 AA.

AC Q9WRU2;  
DT 01-NOV-1999 (TREMblrel. 12, Created)  
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
DT 01-JUN-2000 (TREMblrel. 14, last annotation update)  
DE COMPLEMENT BINDING PROTEIN.  
OS Macaca mulatta rhadinovirus 17577.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 99174001.  
RA Seales R.P., Bergquam E.P., Axthelm M.K., Wong S.W.;  
RT "Sequence and genomic analysis of a rhesus macaque rhadinovirus with  
similarity to Kaposi's sarcoma-associated herpesvirus/human  
herpesvirus 8."  
RL J. Virol. 73:3040-3053(1999).  
DR EMBL; AF083501; AAD21332.1; -.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 8.  
SQ SEQUENCE 645 AA; 71526 MW; 93D8D35ABF61EB2 CRC64;

Query Match  
Best Local Similarity 23.8%; Score 276.5; DB 12; Length 645;





```

Db 52 RCRSGYTTANINATCTGGT--SEPTATCNKSCPNPGEIQNGKVIFFHGQDALYXG 109
OY 90 VKAVYTCNEGVIQLEINREC----DTGWTNDIPICEVYKCLPYAPENGKIVSSAM 144
Db 110 ANIYVVCNEGFLVGRYRCMIGASQAMWSSSPFCEKEXC-----HRPKIENDF 163
OY 145 EPDBREYH-FGQAVRFVNCYKKGIDGDEMHGSDGCFMSEKPKVEICKSPDYINGSPI 203
Db 164 KPKDXYEYNDVAHFECNEGTLVGPISHACAVNNTWTSNMPCELAGCKFPSTVHGYP 223

RESULT 8
P78361
ID P78361 PRELIMINARY; PRT; 347 AA.
AC P78361
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, last sequence update)
DE DECAV-ACCELERATION FACTOR (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN HIPPOCAMPUS;
RA Kumar V.B., Hyung C., Nakra R., Walters M., Sasser T., Bernardo A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U88576; AAB48622.1; -.
DR HSSP: P08603; IHCC.
DR INTERPRO: IPR000436; -.
DR PFM: PF00084; sush1; 4.
FT NON_TER 1
SQ SEQUENCE 347 AA; 37847 MW; 7B123F7F6D48962 CRC64;

Query Match 22.4%; Score 261; DB 4; Length 347;
Best Local Similarity 29.9%; Pred. No. 3.3e-18;
Matches 70; Conservative 41; Mismatches 85; Indels 38; Gaps 12;

OY 1 EDN---ELPRNTELTISWSDQY-PEGTAIKCRGGRY---SLGNVIMCRKGEW 53
Db 58 EFCNRCSEVPTKNSASLQKOPTTQNFVGVTECRGGRREPLSPKTLCLQMLKW 117
OY 54 VALNPLKQCKRPGHGDPTFGTFTLTGNGVFEYGVKAVYTCNEGVIQLEINREC 113
Db 118 --STAVEFCRKKSCPMPEIRNGIDVPGILF--GATIFSCNTGKLFGSTS-SCLT 172
OY 114 DG---WTNDIPICEVYKCLPYAPENGKIVSSAMEPDE-YHFGAVRFVNCYKKG 168
Db 173 SGGSVQWSDPLPCREIYC-----PAPQIDNGIIQGERDHYGRSVIYACNKGFTMIG 227
OY 169 DEEMHC---SDGFMSKEKPKC-----VEISCKSPDYIN-----GSPISOK 206
Db 228 EHSIYCTVANDGEMSGPPPCRGKSLTSKVPPTVHKPTTVNPTVTSPTSOX 281

RESULT 9
O99254
ID O99254 PRELIMINARY; PRT; 679 AA.
AC O99254
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RX MEDLINE: 90229754.
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;

```

```

RT "The murine complement receptor gene family. IV. Alternative splicing
of Cr2 gene transcripts predicts two distinct gene products that share
homologous domains with both human CR2 and CR1."
J. Immunol. 144:3581-3591(1990).
RN [2]
RP SEQUENCE OF 21-367 FROM N.A.
RX MEDLINE: 95105691.
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seya T., Wagner L.M.,
Hollers V.M.;
RT "Mouse complement regulatory protein Crry/p65 uses the specific
mechanisms of both human decay-accelerating factor and membrane
cofactor protein."
J. Exp. Med. 181:151-159(1995).
RL EMBL: U17128; AAA78271.1; -.
DR EMBL: U17123; AAA78271.1; JOINED.
DR EMBL: U17124; AAA78271.1; JOINED.
DR EMBL: U17125; AAA78271.1; JOINED.
DR EMBL: U17126; AAA78271.1; JOINED.
DR EMBL: U17127; AAA78271.1; JOINED.
DR EMBL: M36470; AAA37449.1; -.
DR HSSP: P10998; IVVC.
DR MSD: MGI:88489; Cr2.
DR INTERPRO: IPR000436; -.
DR PFM: PF00084; sush1; 10.
FT NON_TER 1
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 22.4%; Score 260; DB 11; Length 679;
Best Local Similarity 28.4%; Pred. No. 9.5e-18;
Matches 56; Conservative 45; Mismatches 72; Indels 24; Gaps 9;

OY 20 SDQT-YPEGTAIKCRPGYRSLGNVIMCRKGEVALNPLKQCKRPGHGDPTFGTF 78
Db 37 SDKSEFAIGTWEYKCRPGYRFRKSFITTCLETJSKW--SDAQPFCKRRKCPNPEPLHGSV 94
OY 79 TLNGNVEFGVAVYTCNCSYQLEINREC---DTGWTNDIPICEVYKCLPYAP 134
Db 95 HINTG--IEFGSTITTSYCNQGRILIGD-SSATCIVSDNTVMWMDMLCESIPCESPAI 151
OY 135 ENKIVSSAMEPDEYHFGAVRFVNCNSG-----YKIEGDEMHGSD-----GFWSEK 164
Db 152 SNQDFYSSRD---SFFYGNVYTYCHTKNKRKPLVVEKSTYCTSKNOYGIWNSPP 208
OY 185 PKCY-ELISCKSPDYING 200
Db 209 POCIPRYKCPMEIENG 225

RESULT 10
Q28769
ID Q28769 PRELIMINARY; PRT; 522 AA.
AC Q28769
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DE COMPLEMENT RECEPTOR (FRAGMENT).
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Papio.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RA Birmingham D.J., Logar C.M., Shen X.P., Chen W.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L77977; AAA99004.1; -.
DR HSSP: P08603; IHPI.
DR INTERPRO: IPR000436; -.
DR PFM: PF00084; sush1; 7.
FT NON_TER 1
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

```

Query Match 22.0%; Score 256; DB 6; Length 522;  
 Best Local Similarity 31.1%; Pred. No. 1.7e-17;  
 Matches 69; Conservative 38; Mismatches 81; Indels 34; Gaps 14;

3 CNE---ELPPRRNTEILGWSMDQTYEPTGTAIKRCRGRYSLGNVIMVCRKGEWALNPL 59  
 48 CNAPEQLPFPARPTNLTAS----EPYGTLYKYECLPGYKGRKPFSLICLKNVWTSKAD- 102  
 60 RKCQRKPCGHPDTPFGFTFLTGAVNEFYGVKAVYTCNEGTOQLGELNRECDTG---- 115  
 103 -KCTRKSCRNPDPVNGMWHVI--KIDIFGSOJNTSCNKGRLIGS--SATCIIISGTVI 158  
 116 WTNDIPICEVVKC-LPYTAPENCKIVSSAMEPDPREH-FGQAVRFVCSNGS-----YKIE 167  
 159 WDNETPICRIPCGLPPTI-ANGDFISTS---REYFPGSVYTYRNCNLSGSKKLFEIV 213

168 GDEEMHCS--DD--GFWSKEKPKC-VEISCKSPDYINGSPIS 204  
 214 GEPSTYCTSKNDQYGVIMSGPAPCIIIPNKCMPPVENGVLVS 255

RESULT 11  
 ID P87616 PRELIMINARY; PRT; 259 AA.  
 AC P87616;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE 41KB FRAGMENT FROM LEFT END OF GENOME.  
 GN D17L OR C17L.  
 OS Cowpox virus (CEV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GRI-90;  
 RX MEDLINE: 97068532.  
 RA Safonov P.F., Petrov N.A., Riazankina O.I., Totmenin A.V.,  
 RA Shchelkunov S.N., Sandakhchiev L.S.;  
 RT "Genes of a circle of hosts for the cowpox virus.";  
 RL Dokl. Akad. Nauk 349:829-833(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GRI-90;  
 RX MEDLINE: 98229462.  
 RA Shchelkunov S.N., Safonov P.F., Totmenin A.V., Petrov N.A.,  
 RA Ryzankina O.I., Gutorov V.V., Kotwal G.J.;  
 RT "The genomic sequence analysis of the left and right species-specific  
 RT terminal region of a cowpox virus strain reveals unique sequences and  
 RT a cluster of intact ORFs for immunomodulatory and host range  
 RT proteins.";  
 RL Virology 243:432-460(1998).  
 DR EMBL: X94355; CAA64102.1;  
 DR EMBL: Y11842; CAA2367.1;  
 DR HSSP: P10998; IVC.  
 DR INTERPRO: IPR000436;  
 DR PFAM: PF00084; sushi; 4.  
 DR SEQUENCE 259 AA; 28193 MW; 9D1AEF68938859A CRC64;

Query Match 22.0%; Score 255.5; DB 12; Length 259;  
 Best Local Similarity 33.8%; Pred. No. 8.3e-18;  
 Matches 69; Conservative 23; Mismatches 93; Indels 19; Gaps 10;

3 CNEPPRRNTEILGWSMDQTYEPTGTAIKRCRGRY--SLGNVIMVCRKGEWALNPLR 60  
 20 CCLPSPRITMKRGT-VDSHYINGLDIEVLCIPGYKQKMGPIYAKCTGTGTLFN--- 75  
 61 KCQRKPCGHPDTPFGFTFLTGAVNEFYGVKAVYTCNEGTOQLGELNRECDTG---W 116  
 76 QCIRKRCPSRPIDNDGDLIGG---VDFGSSITVSCNSGYHLIGESKSYELGSLGSMV 132

117 TNDIPICEVVKCLPYTAPENCKIVSSAMEPDPREYHFGQAVRFVCSNGYKIEDEEMHCS 176  
 133 NPEAPICESVKQSPSPISNGR--HNGYED--FYDGSVVTYSCNSGSLIGNSGLCS- 187  
 177 DGFWSKEKPKC-VEISCKSPDYING 200  
 188 GGEWS-DPPTCQIVCPHPTISNG 210

RESULT 12  
 ID 090V2 PRELIMINARY; PRT; 559 AA.  
 AC 090V2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE HUMAN CRI mRNA FOR C3b/C4b RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89010527.  
 RA Hourcade D., Miesner D.R., Atkinson J.P., Holers V.M.;  
 RT "Identification of an alternative polyadenylation site in the human  
 RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and  
 RT prediction of a secreted form of complement receptor type 1.";  
 RL J. Exp. Med. 168:1255-1270(1988).  
 DR EMBL: X14362; CAA32541.1;  
 DR INTERPRO: IPR000436;  
 DR INTERPRO: IPR000834;  
 DR PRAM: PF00084; sushi; 8.  
 DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 KW Signal.  
 FT NON\_TER 1  
 FT SIGNAL 1  
 FT CHAIN <1 16 POTENTIAL.  
 FT CHAIN 17 559 POTENTIAL.  
 SO SEQUENCE 559 AA; 61424 MW; DBFPE965CA179D75 CRC64;

Query Match 21.9%; Score 255; DB 4; Length 559;  
 Best Local Similarity 30.3%; Pred. No. 2.4e-17;  
 Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

3 CNE---LPPRRNTEILGWSMDQTYEPTGTAIKRCRGRYSLGNVIMVCRKGEWALNPL 59  
 18 CNAPEWLPFPARPTNLT-----DEFEPIGTIYNTCRGRYSGRPSIICLKNVWTSKAD- 72  
 60 RKCQRKPCGHPDTPFGFTFLTGAVNEFYGVKAVYTCNEGTOQLGELNRECDTG----DTG 115  
 73 -KCTRKSCRNPDPVNGMWHVIK--I0FGSQIKYSCTKGYRLIGS--SATCIIISGTVI 128  
 116 WTNDIPICEVVKC-LPYTAPENCKIVSSAMEPDPREYHFGQAVRFVCSNGS-----YKIEG 168  
 129 WDNETPICRIPCGLPPTI-TNGDFISTNRE---NEHGSVYTYRNCNLSGSKRKFELV 184

169 DEEMHCS--SD--GFWSKEKPKC-VEISCKSPDYINGSPIS 204  
 185 EPSTYCTSKNDQYGVIMSGPAPCIIIPNKCCTPPNVENGILVS 225

RESULT 13  
 ID 016745 PRELIMINARY; PRT; 2039 AA.  
 AC 016745;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN [1]

```
RP SEQUENCE FROM N.A.
RX MEDLINE: 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L17418: AAB60694.1; JOINED.
DR EMBL: L17390: AAB60694.1; JOINED.
DR EMBL: L17391: AAB60694.1; JOINED.
DR EMBL: L17392: AAB60694.1; JOINED.
DR EMBL: L17393: AAB60694.1; JOINED.
DR EMBL: L17394: AAB60694.1; JOINED.
DR EMBL: L17395: AAB60694.1; JOINED.
DR EMBL: L17396: AAB60694.1; JOINED.
DR EMBL: L17397: AAB60694.1; JOINED.
DR EMBL: L17398: AAB60694.1; JOINED.
DR EMBL: L17399: AAB60694.1; JOINED.
DR EMBL: L17400: AAB60694.1; JOINED.
DR EMBL: L17401: AAB60694.1; JOINED.
DR EMBL: L17402: AAB60694.1; JOINED.
DR EMBL: L17403: AAB60694.1; JOINED.
DR EMBL: L17404: AAB60694.1; JOINED.
DR EMBL: L17405: AAB60694.1; JOINED.
DR EMBL: L17406: AAB60694.1; JOINED.
DR EMBL: L17407: AAB60694.1; JOINED.
DR EMBL: L17408: AAB60694.1; JOINED.
DR EMBL: L17409: AAB60694.1; JOINED.
DR EMBL: L17410: AAB60694.1; JOINED.
DR EMBL: L17411: AAB60694.1; JOINED.
DR EMBL: L17412: AAB60694.1; JOINED.
DR EMBL: L17413: AAB60694.1; JOINED.
DR EMBL: L17414: AAB60694.1; JOINED.
DR EMBL: L17415: AAB60694.1; JOINED.
DR EMBL: L17416: AAB60694.1; JOINED.
DR EMBL: L17417: AAB60694.1; JOINED.
DR EMBL: L17418: AAB60694.1; JOINED.
DR EMBL: L17419: AAB60694.1; JOINED.
DR EMBL: L17420: AAB60694.1; JOINED.
DR EMBL: L17421: AAB60694.1; JOINED.
DR EMBL: L17422: AAB60694.1; JOINED.
DR EMBL: L17423: AAB60694.1; JOINED.
DR HSSP: P08603: IHFI.
DR INTERPRO: IPR000436: -.
DR INTERPRO: IPR000834: -.
DR PFM: PFO0084: sushi; 30.
DR PROSITE: PS00087: SOD.CU.ZN.1; UNKNOWN_1.
DR PROSITE: PS00133: CARBOXYPEPT_ZN_2; UNKNOWN_2.
SQ SEQUENCE 2039 AA; 223603 MW; B82FCB11C6B16635 CRC64;
```

```
Query Match 21.9%; Score 255; DB 4; Length 2039;
Best Local Similarity 30.3%; Pred. No. 1.1e-16;
Matches 67; Conservative 39; Mismatches 83; Indels 32; Gaps 13;

QY 3 CNE---LPKRNTELIGSWSQTYPESTQAIYKCRPGYRSIGVNYWCRKGEWALNPL 59
DB 43 CNADEWLPFAFAPNTLT---DEFEPICTLYNYECRPGYSCRPSILTKNSVWIGAD- 97
QY 60 RKCKRPGHGDPTPEGFTLTGNGVEYGVKAYTTCGEGQLLGEINYYRCC---DTDG 115
DB 98 -RCRKRCKRNPDPVYNGVAVIKG--IOPGSOIKYSCRTGKYLIGS--SSAICIIISGDTVI 153
QY 116 WYNDIPICEVVKC-LPYTAPENKIVSSAMEPDREYHFGQAVFVNCSG-----YKIEG 168
DB 154 WDNETPICDRIPICLPPTLI-TNGDFISTNRE---NFHYGSVYTYRKNPFGSGKRVFELVG 209
QY 169 DEEMHC--SDD--GFWSKEKPKC-VEISCKSPDIVNSPIS 204
DB 210 EPTSYCTSNDDGVGWSGPAPOCIIPNKCPTPPNVENGILWS 250

RESULT 14
ID 016744 PRELIMINARY; PRT: 2489 AA.
AC 016744.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE COMPLEMENT RECEPTOR 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RX MEDLINE: 94065175.
RA Vik D.P., Wong W.W.;
RT "Structure of the gene for the F allele of complement receptor type 1
RL J. Immunol. 151:6214-6224(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Vik D.P., Wong W.W.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: L17418: AAB60695.1; JOINED.
DR EMBL: L17390: AAB60695.1; JOINED.
DR EMBL: L17391: AAB60695.1; JOINED.
DR EMBL: L17392: AAB60695.1; JOINED.
DR EMBL: L17393: AAB60695.1; JOINED.
DR EMBL: L17394: AAB60695.1; JOINED.
DR EMBL: L17395: AAB60695.1; JOINED.
DR EMBL: L17396: AAB60695.1; JOINED.
DR EMBL: L17397: AAB60695.1; JOINED.
DR EMBL: L17398: AAB60695.1; JOINED.
DR EMBL: L17399: AAB60695.1; JOINED.
DR EMBL: L17400: AAB60695.1; JOINED.
DR EMBL: L17401: AAB60695.1; JOINED.
DR EMBL: L17402: AAB60695.1; JOINED.
DR EMBL: L17403: AAB60695.1; JOINED.
DR EMBL: L17404: AAB60695.1; JOINED.
DR EMBL: L17405: AAB60695.1; JOINED.
DR EMBL: L17406: AAB60695.1; JOINED.
DR EMBL: L17407: AAB60695.1; JOINED.
DR EMBL: L17408: AAB60695.1; JOINED.
DR EMBL: L17409: AAB60695.1; JOINED.
DR EMBL: L17410: AAB60695.1; JOINED.
DR EMBL: L17411: AAB60695.1; JOINED.
DR EMBL: L17412: AAB60695.1; JOINED.
DR EMBL: L17413: AAB60695.1; JOINED.
DR EMBL: L17414: AAB60695.1; JOINED.
DR EMBL: L17415: AAB60695.1; JOINED.
DR EMBL: L17416: AAB60695.1; JOINED.
DR EMBL: L17417: AAB60695.1; JOINED.
DR EMBL: L17418: AAB60695.1; JOINED.
DR EMBL: L17419: AAB60695.1; JOINED.
DR EMBL: L17420: AAB60695.1; JOINED.
DR EMBL: L17421: AAB60695.1; JOINED.
DR EMBL: L17422: AAB60695.1; JOINED.
DR EMBL: L17423: AAB60695.1; JOINED.
DR EMBL: L17424: AAB60695.1; JOINED.
DR EMBL: L17425: AAB60695.1; JOINED.
DR EMBL: L17426: AAB60695.1; JOINED.
DR EMBL: L17427: AAB60695.1; JOINED.
DR EMBL: L17428: AAB60695.1; JOINED.
DR EMBL: L17429: AAB60695.1; JOINED.
DR EMBL: L17430: AAB60695.1; JOINED.
DR HSSP: P08603: IHFI.
DR INTERPRO: IPR000436: -.
DR INTERPRO: IPR000834: -.
DR INTERPRO: IPR001424: -.
DR PFM: PFO0084: sushi; 37.
DR PROSITE: PS00087: SOD.CU.ZN.1; UNKNOWN_1.
DR PROSITE: PS00133: CARBOXYPEPT_ZN_2; UNKNOWN_3.
SQ SEQUENCE 2489 AA; 272846 MW; CE11B53F2B4FAF6 CRC64;
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:42 ; Search time 49.92 Seconds  
(without alignments)  
181.518 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1 EDCNELPPRRNTEILGSM.....EKSCDNPYPNGDYLRIK 265

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1055	70.5	240	18	Human partial Comp
2	961	64.2	216	18	Clone PRB9PR410 C
3	303	20.2	581	12	Human C4 binding P
4	301.5	20.1	263	20	Vaccinia complemen
5	301	20.1	496	20	Human CRI protein
6	300.5	20.1	543	13	CRI-4 (52S, 53S, 5
7	298.5	19.9	543	13	CRI-4 (318-321 RNP
8	298.5	19.9	543	13	CRI-4 (318-321 RNP
9	298.5	19.9	778	19	Amino acid sequenc
10	298.5	19.9	1930	19	Human complement r
11	298.5	19.9	2039	20	Human C3b/C4b rece
12	298.5	19.9	2317	10	CRI protein. Homo

13	296.5	19.8	543	13	R28560
14	295.5	19.7	543	13	R28570
15	294.5	19.7	543	13	R28550
16	294.5	19.7	543	13	R28553
17	294.5	19.7	543	13	R28565
18	294.5	19.7	543	13	R28568
19	294.5	19.7	543	13	R28569
20	294.5	19.7	543	13	R28571
21	294.5	19.7	2039	12	R1810
22	293.5	19.6	543	13	R28545
23	293.5	19.6	543	13	R28548
24	293	19.6	363	20	W12414
25	293	19.6	363	18	X30918
26	292.5	19.5	543	13	R28555
27	292.5	19.5	543	13	R28558
28	292.5	19.5	543	13	R28563
29	291.5	19.5	543	13	R28549
30	291.5	19.5	543	13	R28551
31	291.5	19.5	543	13	R28562
32	291.5	19.5	543	13	R28566
33	290.5	19.4	254	15	R47154
34	290.5	19.4	254	15	R47155
35	289.5	19.3	263	10	P92003
36	289.5	19.3	543	13	R28556
37	289.5	19.3	543	13	R28554
38	288.5	19.3	543	13	R28552
39	288.5	19.3	543	13	R28561
40	287.5	19.2	543	13	R28554
41	286.5	19.1	543	13	R28544
42	286.5	19.1	543	13	R28546
43	285.5	19.1	543	13	R28543
44	285.5	19.1	543	13	R28559
45	284.5	19.0	263	20	T29858

## ALIGNMENTS

RESULT 1	
W39154	W39154 standard; Protein: 240 AA.
AC	W39154:
XX	27-APR-1998 (first entry)
XX	Human partial Complement factor H protein fragment 1.
DE	Complement factor H: tumour associated antigen; renal cancer;
XX	urogenital cancer; medicament; modulator.
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO9738136-A1.
XX	16-OCT-1997.
PD	
XX	
PF	09-APR-1997; 97WO-US05710.
XX	
PR	06-MAR-1997; 97US-0812481.
PR	09-APR-1996; 96US-0015083.
PR	09-APR-1996; 96US-0630048.
PR	06-MAR-1997; 97US-0038614.
XX	
PA	(BARD-) BARD DIAGNOSTIC SCI INC.
XX	
PI	Enfield DL, Hass GM, Kinders RJ;
XX	WPI, 1997-512742/47.
DR	N-PDB; V02790.
XX	
PT	Treating or screening for cancer, e.g. renal or urogenital cancer -
	by modulating or detecting tumour associated human complement Factor

CRI-4 (114S) analo  
CRI-4 (266-274 KIK  
CRI-4 (64K) analo  
CRI-4 (85R, 87N) a  
CRI-4 (121Q) analo  
CRI-4 (347T, 349Y)  
CRI-4 (369-376 STK  
CRI-4 (364-367 NAA  
Human complement t  
CRI-4 (37Y) analo  
CRI-4 (57V, 59K) a  
Porcine complement  
MCP protein. Unid  
CRI-4 (109N, 110A)  
CRI-4 (117P) analo  
CRI-4 (64K, 65Y) a  
CRI-4 (65T) analo  
CRI-4 (116K) analo  
CRI-4 (318R, 319N)  
Sequence of solubl  
Deduced sequence o  
CRI-4 (94H) analo  
CRI-4 (116K, 117P)  
CRI-4 (78T, 79D) a  
CRI-4 (115T) analo  
CRI-4 (92T, 94H) a  
CRI-4 (35E) analo  
CRI-4 (44T, 47D, 4  
CRI-4 (35E, 37Y) a  
CRI-4 (114-117SKP  
Vaccinia virus SPI

PT H related antigen, or nucleic acid encoding it  
 XX  
 PS Example 6B; Fig 6B; 104pp; English.  
 XX  
 CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone PRB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigens can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.  
 CC  
 SO Sequence 240 AA;  
 Query Match 70.5%; Score 1055; DB 18; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-81;  
 Matches 188; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 78 FTLTGVNFEYGVKAVYTCNEGYLLGEINRREDTGTGNDIPICEVVKCLPYTAPENG 137  
 Db 1 flltgvnfeiygvkavylcnegyqlleinyredtdgtwndipicevkvclpytapeng 60  
 QY 138 KIYSSAMEPDRHYHFGQAVRFVCSGKRIEDDEMHGSDGDFWSEKPKCWEISCKSPDV 197  
 Db 61 kiysamepdrehyhfgavrfvcsngykiegdeemhcsdgdffwsekpkcweisckspdv 120  
 QY 198 INSSPIQKTIYKENERFYCKNMGYYSERGDVAVCTESGWRPLPSCSEKSCNDPYTPNG 257  
 Db 121 inspslqkilykenerfyckmgyeysergdavctesgwrplpsceekscndpytpng 180  
 QY 258 DYSPLRIK 265  
 Db 181 dysplrik 188  
 RESULT 2  
 ID W39155 standard; Protein; 216 AA.  
 XX  
 AC W39155;  
 XX  
 DT 27-APR-1998 (first entry)  
 XX  
 DE Clone PRB9FH410 CFH related protein fragment.  
 XX  
 KW Complement factor H; tumour associated antigen; renal cancer;  
 KW urogenital cancer; medicament; modulator.  
 OS  
 OS Synthetic.  
 PN WO9738136-A1.  
 PD 16-OCT-1997.  
 PF 09-APR-1997; 97WO-US05710.  
 PR 06-MAR-1997; 97US-0812481.  
 PR 09-APR-1996; 96US-0015083.  
 PR 09-APR-1996; 96US-0630048.  
 PR 06-MAR-1997; 97US-0038614.  
 PA (BARD-) BARD DIAGNOSTIC SCI INC.  
 PI Enfield DL, Hass GM, Kinders RJ;  
 DR WPI: 1997-512742/47.  
 DR N-PSDB: V02791.  
 PT Treating or screening for cancer, e.g. renal or urogenital cancer -  
 PT by modulating or detecting tumour associated human Complement Factor  
 PT H related antigen, or nucleic acid encoding it

XX  
 PS Example 6B; Fig 6B; 104pp; English.  
 XX  
 CC This partial protein is found in clone PRB9FH410 and represents a  
 CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.  
 CC  
 SO Sequence 216 AA;  
 Query Match 64.2%; Score 961; DB 18; Length 216;  
 Best Local Similarity 99.4%; Pred. No. 2.2e-73;  
 Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 95 TCNNGYQLLGINRREDTGTGNDIPICEVVKCLPYTAPENGKIYSSAMEPDRHYHFGQ 154  
 Db 1 tcnegyqlleinyredtdgtwndipicevkvclpytapengkiysamepdrehyhfgq 60  
 QY 155 AVRFVCSGKRIEDDEMHGSDGDFWSEKPKCWEISCKSPDVINGSPISQKIYKENER 214  
 Db 61 avrfvcsngykiegdeemhcsdgdffwsekpkcweisckspdvingspsqkilykener 120  
 QY 215 FQYCKNMGYYSERGDVAVCTESGWRPLPSCSEKSCNDPYTPNGDYSPLRIK 265  
 Db 121 fqyckmgyeysergdavctesgwrplpsceekscndpytpngdysplrik 171  
 RESULT 3  
 ID R13490 standard; Protein; 581 AA.  
 XX  
 AC R13490;  
 XX  
 DT 30-OCT-1991 (first entry)  
 XX  
 DE Human C4 binding protein.  
 XX  
 KW C4bp; monomer; complement protein; pJOD.c4bp.3; SCR;  
 KW short consensus repeat.  
 OS  
 OS Homo sapiens.  
 FH Key  
 FH Peptide  
 FT 1..32  
 FT /label= signal\_peptide  
 FT 33..581  
 FT Protein  
 FT /label= C4bp  
 FT 33..93  
 FT /label= SCR8  
 FT 94..155  
 FT /label= SCR7  
 FT 156..219  
 FT /label= SCR6  
 FT 220..279  
 FT /label= SCR5  
 FT 280..345  
 FT /label= SCR4  
 FT 346..406  
 FT /label= SCR3  
 FT 407..464  
 FT /label= SCR2  
 FT 465..523  
 FT /label= SCR1  
 FT 524..581  
 FT /label= C4bp-core  
 FT /note= "responsible for multimer assembly"  
 FT 34..80  
 FT /note= "intradomain"  
 FT 65..92  
 FT Disulfide-bond  
 FT Disulfide-bond



FT		/note= "Intracdomain"
xx		
PN	WO9111461-A.	
xx		
PD	08-AUG-1991.	
xx		
PF	28-JAN-1991;	91WO-US00567.
xx		
PR	26-JAN-1990;	90US-0470888.
xx		
PA	(BIOG-) BIOGEN INC.	
xx		
Pasek MP,	Minkler G,	Liu TR:
xx		
PI		
DR	WPI: 1991-252613/34.	
xx		
N-PSDB:	Q13242.	
xx		
PT	New C4 binding protein fusion proteins and DNA encoding them -	
xx	comprise assemblies of C4bp monomers linked to functional moiety,	
PT	e.g. AZT, useful as delivery vehicles in diagnosis and therapy	
xx		
PS	Example 1; Fig 1; 105pp; English.	
xx		
CC	This sequence was deduced from human hepatocyte (hep G2) cDNA	
CC	obtained following PCR amplification. The protein is a monomer	
CC	containing 8 SCRs. Each SCR forms a looped domain due to the	
CC	presence of two intradomain disulphide bonds (only the disulphide	
CC	bonds of SCR8 are labelled in the Features Table). Within each SCR,	
CC	the first cysteine residue binds with the third and the second	
CC	cysteine residue binds with the fourth. This secondary structure is	
CC	responsible for the conformational flexibility of the C4bp monomer.	
CC	The invention covers fusion proteins in which the monomer sequence,	
CC	or a specified subfragment of it, e.g. having 5, 4, 3 or 1 SCR(s)	
CC	is fused to the C-terminal of a protein such as a viral receptor,	
CC	cell ligand, a bacterial, viral or parasitic immunogen, enzyme,	
CC	cytokine, toxin, etc. See also Q13243-51.	
xx		
SQ	Sequence	581 AA:
	Query Match	20.2%; Score 303; DB 12; Length 581;
	Best Local Similarity	30.7%; Pred. No. 1,2e-17;
	Matches	80; Conservative 42; Mismatches 101; Indels 38; Gaps 16;
QY	20 SNOTPREGQAIIKCRPG-RSLGNVIMYCKR-GEVVALNPLRKCKRGPHGDPPFGT	77
	:::	
Db	51 tetrktfctllkytlcplgyvzshstqtlclnsdgewv-yntf-cltykcrlrhpge-----	102
QY	78 FTLVTGGNV-----FEYGVAAYTCNGDGYOLLGEINRECDTD---GWTNDIPICEVKC	128
Db	103 --lmgqvteikcdlsifgsdfeiscsegfllystusr-cevqdrgyvwshlpipqeivvc	159
QY	129 LPTAPENCKIVSSAMEPDREYHFGQAVRFVCNSGGXIKIGDEEMHCSDS---GFWSKER	184
Db	160 kppddlngdrhsge-----enfaygfsvtyscdprfslghasistcvtenetlgwvrpsp	215
QY	185 PKCVELSCSKSPVINGSPISQ-KIITYKENEROYCKNMGEYSERDAY--C-TESGWPR	240
Db	216 pteckltctckpdsvisgwmsgfgprlynykdltvivfkcgkfvl--tggsvlhcdadsdkmp	273
QY	241 L-PSCEEKSCDN-PYIPNGDY	259
Db	274 sppacepcnscnlpdiphasw	294
RESULT	4	
ID	y29859	
xx	y29859 standard; Protein; 263 AA.	
xx		
AC	y29859;	
xx		
DT	16-NOV-1999	(first entry)
xx		

DE		Vaccinia complement control protein sequence.	
XX			
KW		Vaccinia virus; hyperacute inhibitor of complement enzyme; SPICE;	
KM		fusion protein; hyperacute rejection; xenograft; inflammation;	
KW		post-ischaemic reperfusion injury; malignancy; autoimmune disease;	
KM		immune system disorder; neurodegeneration; infection; gene therapy;	
KW		blood additive; extracorporeal circulation system.	
XX			
OS		Vaccinia virus.	
XX		Synthetic.	
PN		MO9944625-A1.	
PD		10-SEP-1999.	
XX			
PF		02-MAR-1999; 99WO-USO4635.	
XX			
PR		03-MAR-1998; 98US-0076821.	
PA		(UYJO ) UNIV JOHNS HOPKINS.	
PA		(UYPI-) UNIV PITTSBURGH.	
PI		Rosengard AM, Ahearn JM;	
DR		WPI: 1999-550981/46.	
N-PSDB:		Z21091.	
PT		New smallpox inhibitor of complement enzyme protein, used to treat	
XX		complement-mediated disease, particularly hyperacute rejection -	
PS		Claim 15; Fig 1; 88pp; English.	
XX			
CC		The present invention describes the Vaccinia virus smallpox inhibitor of	
CC		complement enzymes (SPICE) protein. SPICE is an inhibitor of complement	
CC		activation, and so can be used to treat or prevent complement-mediated	
CC		disorders, especially hyperacute rejection, inflammation or post-	
CC		ischaemic reperfusion injury, malignancies, autoimmune diseases,	
CC		immune system disorders, neurodegeneration and infections. Hyperacute	
CC		rejection may also be prevented by treating the graft with SPICE before	
CC		transplanting it or by using a xenograft that has been transformed to	
CC		express SPICE from a gene therapy vector. SPICE is also useful as	
CC		additive to blood, e.g. in an extracorporeal circulation system (coated	
CC		on tubing) or in storage, also for studying complement activation.	
CC		Transgenic animals that express SPICE are used as sources of xenografts.	
CC		The present sequence represents a vaccinia complement control protein	
CC		(VCP) encoded by the specifically claimed mutated VCP nucleotide	
CC		sequence, having a silent T to A transversion at nucleotide position	
CC		number 267.	
CC			
SO		Sequence 263 AA;	
	Query Match	20.1%; Score 301.5; DB 20; Length 263;	
	Best Local Similarity	33.3%; Pred. No. 6; 2e-18;	
	Matches 78; Conservative	29; Mismatches 106; Indels 21; Gaps 12.	
OY	20	SPDTPEEGQATKYCKPGGR--SLGVNIMWCKGGEVALNPRLRKOKRPGCHBGPDPFGT 77	
Db	40	aananyigdtietlylcpjyrkykqmpriyakctqgtwtlfn---qclktcpspridinq 96	
OY	78	FLLTGNGVEYGCAVAYTYNEGYYOLGET-NYRECDTDG--WTNDIPICEVKKCLPYTA 133	
Db	97	ldlg---vdfgsstlyscnsghyllgeakscyalsctsgsmwnpeaplcswkcqspps 153	
OY	134	PENGKIVSSAMERDREYHNHGOAVRFNCNSGYKIEDGBEHNCSDDGWMSKEKRCVPDISK 193	
Db	154	isngr-hngyed-fyldgsvvlyscnsysllgnsgylcs-ggews-dpdcqlvkcp 207	
OY	194	SPDVINGSPIISO-KIITYENERFOYKCNNAGYEYSEGDJAVCTESG-WRP-LPSC 244	
Db	208	hptlingylasgfkrisyndnvdckfygyrksagssstcspagnwkpelpkc 261	

RESULT 5  
 Y55752 standard: Protein: 496 AA.  
 AC Y55752;  
 DT 22-FEB-2000 (first entry)  
 DE Human CRI protein LHR-A SCR fragment.  
 XX  
 XX C3B/C4B receptor; CRI protein; cell-surface protein; erythrocyte; human;  
 KM complement regulatory activity; complement pathway enzyme; tissue damage;  
 KM reperfusion injury; Arthus reaction; myocardial infarct; inflammation;  
 KM heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;  
 KM short consensus repeat.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US981481-A.  
 PD 09-NOV-1999.  
 XX  
 PF 06-JUN-1995: 95US-0470652.  
 XX  
 PR 03-APR-1989: 89US-0332865.  
 PR 06-DEC-1974: 74US-0350238.  
 PR 24-FEB-1993: 93US-0026134.  
 PR 01-APR-1988: 88US-0176532.  
 XX  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.  
 XX  
 PI Concino MF, Wong MW, Makrides SC, Klickstein LB, Fearon DT, IP SH;  
 PI Marsh HC, Carson GR;  
 XX  
 DR WPI: 1999-633357/54.  
 XX  
 PT A human C3B/C4B receptor (CRI) protein having antinflammatory and  
 PT cardiant activity -  
 XX  
 PS Disclosure: Fig 10; 87pp: English.  
 XX  
 XX The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI  
 CC protein or fragment is expressed as a cell-surface protein on the surface  
 CC of a non-human cell and exhibits a complement regulatory activity of full  
 CC -length human CRI as expressed on erythrocytes. The CRI function in vivo  
 CC may be mediated through the inhibition of complement pathway enzymes. The  
 CC soluble CRI protein exhibits a complement regulatory activity, and this  
 CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and  
 CC neutrophil mediated tissue damage, and reduce myocardial infarct size,  
 CC and inflammation. The CRI protein and its fragments can also be used in  
 CC the treatment of conditions which involve unwanted complement activity,  
 CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,  
 CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti-  
 CC -CRI antibodies are used in assays, and diagnostics. The present sequence  
 CC represents the short consensus repeat (SCR) fragments of human CRI  
 CC protein long homologous repeat (LHR)-A sequence.  
 XX  
 SQ Sequence 496 AA:

Query Match 20.1%; Score 301; DB 20; Length 496;  
 Best Local Similarity 27.5%; Pred. No. 1.5e-17;  
 Matches 88; Conservative 48; Mismatches 110; Indels 74; Gaps 19;

QY 3 CNE---LPPRNTETLLTGSWSQDTTPEGTAIYKCRPGYGRSGNIMWCRKEB-WVALNP 58  
 DB 48 cnapewlpfarptltt---defefpigtlylncycrpygsrpfslclcknswtgaad 103  
 QY 59 LKRCCKRPGCGHGDTPFGFTLTGGNVPEYGVKAVYTCNCEGYQLGEIYRRC---PTD 114  
 DB 104 --rcrkscmpdpvnmvnhvikg--igtgsqlyksckkyrligs--ssatclisgdtlv 158

QY 115 GWINDIPICEYKVC-LPVLAPENGKIVSSAMEPDREYHFGQAVRFVNSG-----YKIE 167  
 DB 159 iwdnetpdcdrtpcglpplcl-tngdfistnre---nfhygsvlyrcmpnsggkvtfelv 214  
 QY 168 GDEEMHC--SDD--GFWSEKPKC-VEISCKSPDYVINGSPIS-QKTIYKENERQYCNM 221  
 DB 215 gepaiyctsnddgyglwsgpapgclpncctphvnengllvsdnrslfslnevevtfcqp 274  
 QY 222 GYEYSEBGDAVCTE-SCWRP-LPSCDE-----KSCDN 251  
 DB 275 gfvmkgyrrvkcqqlnkwepeipscsvvcqppdvlnaertqrkdnfsgygcfcqp 334  
 QY 252 PY-----IPNGDYSP 261  
 DB 335 gydlrgaasmrcipgqdwsp 354

RESULT 6  
 R28557  
 ID R28557 standard: peptide: 543 AA.  
 AC R28557;  
 XX  
 DT 19-MAR-1993 (first entry)  
 XX  
 DE CRI-4 (99H, 103E) analogue.  
 XX  
 KM short consensus repeat; regulator of complement activation;  
 KM C3b binding; C4b binding; human complement type 1 receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Region 1..60  
 FT /label= SCR-1  
 FT Region 61..122  
 FT /label= SCR-2  
 FT Region 451..510  
 FT /label= SCR-8  
 FT Region 511..543  
 FT /label= SCR-9  
 FT /note= "TRUNCATED"  
 FT Misc-difference 99  
 FT /note= "Ser substituted by His (SCR-9)"  
 FT FT Misc-difference 103  
 FT /note= "Thr substituted by Glu (SCR-9)"  
 PN EP512733-A.  
 XX  
 PD 11-NOV-1992.  
 XX  
 PF 28-APR-1992: 92EP-0303826.  
 XX  
 PR 03-MAY-1991: 91US-0695514.  
 XX  
 PA (UNTW ) UNIV WASHINGTON.  
 XX  
 PI Atkinson JP, Hourcade D, Kiyeh M;  
 XX  
 DR WPI: 1992-375009/46.  
 XX  
 PT Complement activity regulator protein analogues - useful for  
 PT treating autoimmune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 XX  
 PS Claim 11; Fig 2 and R11810; 23pp: English.  
 XX  
 CC The cDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of



XX		short consensus repeat; regulator of complement activation;
KW	C3b binding; C4b binding; human complement type I receptor.	
XN		
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	Region	/label= SCR-1 1..60
FT	Region	/label= SCR-1 61..122
FT	Region	/label= SCR-2 451..510
FT	Region	/label= SCR-8 511..543
FT	Misc-difference	/note= "TRUNCATED"
FT	Misc-difference	/note= "Asp substituted by Arg"
FT	Misc-difference	/note= "Asp substituted by Asn"
FT	Misc-difference	/note= "Phe substituted by Pro"
FT	Misc-difference	/note= "Met substituted by Pro"
PX	EP5I2733-A.	
PN	11-NOV-1992.	
PD	28-APR-1992:	92EP-0303826.
XX	PR	03-MAR-1991: 91US-0695514.
XX	(UNIM ) UNIT V WASHINGTON.	
PA	Atkinson JP, Hourcade D, Krych M; WPI: 1992-375009/46.	
DR	Complement activity regulator protein analogues - useful for treating auto-immune diseases, to suppress transplant rejection, for diagnosis etc.	
PS	Example 8; Page 18 and R11810; 23pp; English.	
CC	The cDNA clone designated CR1-4 was described in J Exp Med.(1988) C3C 168:1255-1270. It encodes the first 8 and a half amino terminal SCRs of CR1. The invention concerns analogues of "regulator of complement activation" proteins or truncated, hybrid or recombinant forms of them. CR1-4 is a preferred truncated form and a number of specified substitution variants of it are disclosed in which certain positions in SCR-5-6 are substituted by amino acids from the corresponding positions in SCRs which are involved in C3b- and C4b-binding. The substitution variant given here has increased C4b- binding. The specification does not contain the CR1-4 sequence; the sequence given here was constructed from the full-length CR1 amino acid sequence having GENSEQ accession number R11810 and descriptions in the disclosure.	
XX	Sequence	543 AA:  Query Match            19.9%; Score 298.5; DB 13; Length 543; Best Local Similarity   26.5%; Pred.No.2,7e-17; Matches      88; Conservative     44; Mismatches   105; Indels       95; Gaps          18
OY	3 CNE---LPPRNTEILLGSWDOTYPEGTAOATIKCRPGYSLGNIWMVRKGEVALNEL	59
Dd	2 cnapewldfarplntt---defefrpfygllyncrcprysgypfsfclknswtgakd-	56
OY	RKCCKRCPCGHDPDIPFGFETLLTGWNVEEGLAAVTTCNGCYOLLCEINTRYEC---	DTGG 115
Dd	-tcrtkscrpdpdvgnmhwltky--lqfgsqdkyscktyrlilgs-saatcilsdgdlv1	112

QY	116	WTNIPICDEVK-LPTVAENCKIYSSAMEPDRYHFQAVRFEVCGSG-----YRIEG	168
Db	113	wdnetpictdrpoglpdlti-tngdfisthre--nfhygsvvtylcmpgsgkrvfelvg	168
QY	169	DEEMHC--SDD--GFMSKEPKC-VEISCKSPDVINGSPIS-QKTIYEN-----	212
Db	169	epsilystcnddqygiwsgpapqcllpnkctprnvengllvdsndrlsflsinevfeircqv	228
QY	213	-----ERFQ-----YKCNMG	222
Db	229	fvmkgprrrvkcgaalnkwepelpscsvvcqppdvhaertqdrkdnfsgpgvevlyscpg	288
QY	223	YEYBERGDVACTESG-WRP-LPSCEEKSCDNP	252
Db	289	ydlrgaasmrcptpgdwspapcvcckscrrp	320
RESULT	9		
W73147			
ID	W73147	standard; protein; 778 AA.	
XX	W73147;		
XX			
DT	29-JAN-1999	(first entry)	
XX			
DE		Amino acid sequence of the soluble complement receptor 1 (SCR1).	
XX			
KW		Human; soluble complement receptor 1; SCR1; T-cell; B-cell;	
KW		mediated immune response; inhibition; tissue rejection; gene therapy;	
KW		dystrophin; inflammatory response; interferon-gamma secretory response	
KW		autoimmune response; neurological response; Alzheimer's disease;	
KW		Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;	
KW		rheumatoid arthritis; myasthenia gravis; epidermis bullosa;	
KW		Hashimoto's disease.	
XX			
OS		Homo sapiens.	
XX			
PN		MO9845430-A1.	
XX			
PD		15-OCT-1998.	
XX			
PE		06-APR-1998; 98MO-GB01012.	
PR		05-APR-1997; 97GB-0006950.	
XX			
PA		(ANNE/) ANNENKOV A.	
XX		(CHER/) CHERNAJOVSKY Y.	
XX			
PL		Annenkov A, Chernajovsky Y;	
DR		WPI; 1998-568350/48.	
XX			
PT		Fragment of soluble human complement receptor 1 - useful for	
PT		treating T-cell or B-cell mediated immune responses e.g.	
PT		inflammatory responses such as rheumatoid arthritis	
XX			
PS		Disclosure; Fig 1; 54pp; English.	
XX			
CC		This is an amino acid sequence of the human soluble complement	
CC		receptor 1 (SCR1), useful in the treatment of T-cell or B-cell	
CC		mediated immune responses. It is used to inhibit a T-cell or	
CC		B-cell-mediated immune response to prevent immune response-mediated	
CC		tissue rejection and destruction or clearance or inactivation of an	
CC		expressed protein especially from cells that have been treated by gene	
CC		therapy to express the protein, e.g. dystrophin. The protein can also	
CC		be used to inhibit a T-cell or B-cell-mediated inflammatory response,	
CC		an interferon-gamma secretory response, autoimmune response or	
CC		neurological response, e.g. Alzheimer's or Parkinson's disease or	
CC		multiple sclerosis. Also the protein can be used to treat systemic	
CC		lupus erythematosus, rheumatoid arthritis, myasthenia gravis,	
CC		epidermis bullosa or Hashimoto's disease.	
XX			



PR 06-DEC-1974; 74US-0350238.  
 PR 24-FEB-1993; 93US-0026134.  
 PR 01-APR-1988; 88US-0176532.  
 XX  
 XX (UYJO ) UNIV JOHNS HOPKINS.  
 PA (BGHM ) BRIGHAM & WOMEN'S HOSPITAL.  
 PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.  
 XX  
 PI Concino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;  
 PI Marsh HC, Carson GR;  
 DR MPI; 1999-633357/54.  
 DR N-PSDB; 238150.  
 XX  
 PT A human C3B/C4B receptor (CRI) protein having antiinflammatory and  
 PT cardiant activity -  
 XX  
 XX Disclosure; Fig 1A-P; 87pp; English.  
 XX  
 CC The invention relates to a human C3B/C4B receptor (CRI) protein. The CRI  
 CC protein or fragment is expressed as a cell-surface protein on the surface  
 CC of a non-human cell and exhibits a complement regulatory activity of full  
 CC -length human CRI as expressed on erythrocytes. The CRI function in vivo  
 CC may be mediated through the inhibition of complement pathway enzymes. The  
 CC soluble CRI protein exhibits a complement regulatory activity, and this  
 CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and  
 CC neutrophil mediated tissue damage, and reduce myocardial infarct size,  
 CC and inflammation. The CRI protein and its fragments can also be used in  
 CC the treatment of conditions which involve unwanted complement activity,  
 CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,  
 CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti  
 CC -CRI antibodies are used in assays, and diagnostics. The present sequence  
 CC represents the human CRI protein.  
 XX  
 SQ Sequence 2039 AA:  
 Query Match 19.9%; Score 298.5; DB 20; Length 2039;  
 Best Local Similarity 27.0%; Pred. No. 1.4e-16;  
 Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;  
 QY 3 CNE---LPPRRNTTEILGSMDSQDTPEGTQAIYKCRPGYRSLGNVIMVCKRGWALNPL 59  
 DB 43 chapeawlptarplntl----defefpigtlylneyecrpyysgrpfslclknswtgakd- 97  
 QY 60 RCOGRPCGHPGPTPGTFTLTGNGNFEYGVKAVYTCNCEYOLGELINREC-----DTDG 115  
 DB 98 -rrrrscnppdpvngmwhvlykg--lqfsgqikysctkyrllgs--saatcllsgdtvi 153  
 QY 116 WTNDIPICEVVKC-LPVTAPENCKIYSSAMEPDRREYHFGQAVRFVNCNSG-----YKIEG 168  
 DB 154 wdneprldiripglpftl-tngdfistnre--nhysgvlyrcnpsggrkvfelvg 209  
 QY 169 DEEMHC--SDD--GFWSKREKPC-VEISCKSPVYINGSPIS-OKIYKENERQYVCNMG 222  
 DB 210 epslyctsnddvgylwsgpdpqcllpnkctpvengllvsdmsrlsfsjnevevficpqg 269  
 QY 223 YEYSESGDAVCTE-SGMRP-LPSCDE-----KSCDNP 252  
 DB 270 fymkpgprvkqgalnkweelpscsrvcqppdvlnhaertgrdkdnfsgqevfyscepg 329  
 QY 253 Y-----IPNGDYSP 261  
 DB 330 ydlrgaasmrctpgqdwsp 348  
 Db  
 RESULT 12  
 P92219  
 ID P92219 standard; protein; 2317 AA.  
 XX  
 AC P92219;  
 XX  
 DT 22-FEB-1990 (first entry)

XX  
 DE CRI protein.  
 XX  
 XX Complement; cofactor.  
 KM  
 XX Homo sapiens (human).  
 OS  
 XX  
 FH Key location/Qualifiers  
 FT Peptide 10..50  
 FT /label= signal\_peptide  
 XX  
 PN W08909220-A.  
 XX  
 XX  
 PD 05-OCT-1989.  
 XX  
 XX  
 PE 31-MAR-1989; 89MO-US01358.  
 XX  
 PR 01-APR-1988; 88US-0176532.  
 XX  
 XX (TCEL ) T CELL SCIENCES INC.  
 PA (UYJO ) THE JOHNS HOPKINS UNIVERSITY.  
 PA (BRIG ) THE BRIGHAM AND WOMEN'S HOSPITAL.  
 XX  
 PI Fearon DT, Klickstein LB, Wong W, Carson G, Concino MF, Makrides SC;  
 DR MPI; 1989-309498/42.  
 DR N-PSDB; N91477.  
 XX  
 PT New nucleic acid sequences encoding new CRI protein - and its fragment,  
 PT for diagnosis and control of complement-related immune defects,  
 PT inflammation, myocardial infarct, etc  
 XX  
 XX Claim 1; Fig 1; 191pp; English.  
 PS  
 XX  
 CC This is full-length CRI protein, and shortened forms are new, lacking  
 CC the transmembrane region. The proteins and fragments bind C3b and/or  
 CC C4b, have cofactor activity and inhibit C3 and C5 convertase activity.  
 CC In the sequence, x=untranslated region. This has 7 short consensus  
 CC repeats (CR) of 4 long direct homologous repeats (LHR) (see fig. 10 in  
 CC patent). Deletion mutants, eg LHR-A for binding of C4b, and LHR-C  
 CC for C3b binding, be made, and LHR-B and -C can detect CR2 sequences.  
 CC They are useful in diagnosing and treating immune disorders, and prevent  
 CC perfusion injury.  
 XX  
 SQ Sequence 2317 AA:  
 Query Match 19.9%; Score 298.5; DB 10; Length 2317;  
 Best Local Similarity 27.0%; Pred. No. 1.6e-16;  
 Matches 86; Conservative 49; Mismatches 111; Indels 73; Gaps 18;  
 QY 3 CNE---LPPRRNTTEILGSMDSQDTPEGTQAIYKCRPGYRSLGNVIMVCKRGWALNPL 59  
 DB 52 chapeawlptarplntl----defefpigtlylneyecrpyysgrpfslclknswtgakd- 106  
 QY 60 RCOGRPCGHPGPTPGTFTLTGNGNFEYGVKAVYTCNCEYOLGELINREC-----DTDG 115  
 DB 107 -rrrrscnppdpvngmwhvlykg--lqfsgqikysctkyrllgs--saatcllsgdtvi 162  
 QY 116 WTNDIPICEVVKC-LPVTAPENCKIYSSAMEPDRREYHFGQAVRFVNCNSG-----YKIEG 168  
 DB 163 wdneprldiripglpftl-tngdfistnre--nhysgvlyrcnpsggrkvfelvg 218  
 QY 169 DEEMHC--SDD--GFWSKREKPC-VEISCKSPVYINGSPIS-OKIYKENERQYVCNMG 222  
 DB 219 epslyctsnddvgylwsgpdpqcllpnkctpvengllvsdmsrlsfsjnevevficpqg 278  
 QY 223 YEYSESGDAVCTE-SGMRP-LPSCDE-----KSCDNP 252  
 DB 279 fymkpgprvkqgalnkweelpscsrvcqppdvlnhaertgrdkdnfsgqevfyscepg 338  
 QY 253 Y-----IPNGDYSP 261  
 DB 279 fymkpgprvkqgalnkweelpscsrvcqppdvlnhaertgrdkdnfsgqevfyscepg 338



168:1255-1270. It encodes the first 8 and a half amino terminal SCRs of CRI. The invention concerns analogues of "regulator of complement activation" proteins or truncated, hybrid or recombinant forms of them. CRI-4 is a preferred truncated form and a number of specified substitution variants of it are disclosed in which certain positions in SCR-5-6 are substituted by amino acids from the corresponding positions in SCRs which are involved in C3b- and C4b-binding. The substitution variant given here has increased C3b-binding. The specification does not contain the CRI-4 sequence; the sequence given here was constructed from the full-length CRI amino acid sequence having GENESEQ accession number R11810 and descriptions in the disclosure.

Sequence 543 AA:

Query Match 19.7%; Score 295.5; DB 13; Length 543;  
Best Local Similarity 26.0%; Pred. No. 4.8e-17;  
Matches 86; Conservative 44; Mismatches 106; Indels 95; Gaps 17;

```

QY 3 CNE---LPPRRMTETITGSMSPDQTYEGTQATYKCRPGYRSLGNVIMCRKGWALNPL 59
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2 cnapewlrfarptnlt-----defefipigtylnyecrpysgrpfslclknswtgcakd- 56
QY 60 KCKQKRPCGHEGDPFGFTLTGNAVFEYGVKAVYTCNEGQQLGEINRYREC---DTDG 115
   || || || || || || || || || || || || || || || || || || || || || ||
Db 57 -rrrkscskppdpvngmhwikg--lqfsgqikyscktyrllgs--ssatclisgdlvi 112
QY 116 WNDPIPCVVKC-LPVYAPENGKIVSSAMEPDREYHFGQAVRFVNCNG-----YKIEG 168
   || || || || || || || || || || || || || || || || || || || || || ||
Db 113 wdneplcdrlpcglppll-tngdfisture---nfhygsvvlyrcnpgsggrkvfelvg 168
QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDYINGSPIS--QKRIYEN----- 205
   || || || || || || || || || || || || || || || || || || || || || ||
Db 169 epslyctsnddgvjwsppapcqlpnkctprnvengllvsdnrlfslnevefrcqpv 228
QY 206 -----KTIYKENERFQ-----YKCNMG 222
   || || || || || || || || || || || || || || || || || || || || || ||
Db 229 fvmkgprrvkcqalnkweplscsrvcqpppdlvhaerlqrdknfispqgevfyscepg 288
QY 223 YEYERGDVCTESG-WRP-LPSCSEKSCDN 251
   || || || || || || || || || || || || || || || || || || || || || ||
Db 289 ydlrgaasmrctpgqdwspaaptcevkscdd 319

```

RESULT 15  
R28550 ID R28550 standard; peptide; 543 AA.  
XX R28550;

DT 19-MAR-1993 (first entry)  
XX  
DE CRI-4 (64K) analogue.  
XX  
KW short consensus repeat; regulator of complement activation;  
XX C3b binding; C4b binding; human complement type 1 receptor.  
XX  
OS Homo sapiens.  
XX

XX Key Location/Qualifiers  
FH Region 1..60  
FT /label= SCR-1  
FT Region 61..122  
FT /label= SCR-2  
FT Region 451..510  
FT /label= SCR-8  
FT Region 511..543  
FT /label= SCR-9  
FT /note= "TRUNCATED"  
XX Misc-difference 64  
/note= "Arg substituted by Lys (SCR-9)"

PN EP512733-A.  
XX  
XX 11-NOV-1992.  
PD  
PF 28-APR-1992; 92EP-0303826.  
XX  
XX 03-MAY-1991; 91US-0695514.  
PR  
XX (UNITW ) UNIV WASHINGTON.  
XX  
XX Atkinson JP, Hourcade D, Krych M;  
PI WPI; 1992-375009/46.  
XX  
XX  
XX Complement activity regulator protein analogues - useful for  
PT treating autoimmune diseases, to suppress transplant rejection,  
PT for diagnosis etc.  
PS  
XX Claim 11; Fig 2 and R11810; 23pp; English.

The cDNA clone designated CRI-4 was described in J Exp Med.(1988)  
CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
CC SCRs of CRI. The invention concerns analogues of "regulator of  
CC complement activation" proteins or truncated, hybrid or recombinant  
CC forms of them. CRI-4 is a preferred truncated form and a number of  
CC specified substitution variants of it are claimed in which certain  
CC positions in SCR-2 which have been identified as important for the  
CC degree of C3b- and C4b-binding are substituted by amino acids from  
CC the corresponding positions in SCR-9. The specification does not  
CC contain the CRI-4 sequence; the sequence given here was constructed  
CC from the full-length CRI amino acid sequence having GENESEQ  
CC accession number R11810 and descriptions in the disclosure.

Sequence 543 AA:

Query Match 19.7%; Score 294.5; DB 13; Length 543;  
Best Local Similarity 26.3%; Pred. No. 5.8e-17;  
Matches 87; Conservative 45; Mismatches 104; Indels 95; Gaps 18;

```

QY 3 CNE---LPPRRMTETITGSMSPDQTYEGTQATYKCRPGYRSLGNVIMCRKGWALNPL 59
   || || || || || || || || || || || || || || || || || || || || || ||
Db 2 cnapewlrfarptnlt-----defefipigtylnyecrpysgrpfslclknswtgcakd- 56
QY 60 KCKQKRPCGHEGDPFGFTLTGNAVFEYGVKAVYTCNEGQQLGEINRYREC---DTDG 115
   || || || || || || || || || || || || || || || || || || || || || ||
Db 57 -rrrkscskppdpvngmhwikg--lqfsgqikyscktyrllgs--ssatclisgdlvi 112
QY 116 WNDPIPCVVKC-LPVYAPENGKIVSSAMEPDREYHFGQAVRFVNCNG-----YKIEG 168
   || || || || || || || || || || || || || || || || || || || || || ||
Db 113 wdneplcdrlpcglppll-tngdfisture---nfhygsvvlyrcnpgsggrkvfelvg 168
QY 169 DEEMHC--SD--GFWSKEKPKC-VEISCKSPDYINGSPIS--QKRIYEN----- 212
   || || || || || || || || || || || || || || || || || || || || || ||
Db 169 epslyctsnddgvjwsppapcqlpnkctprnvengllvsdnrlfslnevefrcqpv 228
QY 213 -----ERFQ-----YKCNMG 222
   || || || || || || || || || || || || || || || || || || || || || ||
Db 229 fvmkgprrvkcqalnkweplscsrvcqpppdlvhaerlqrdknfispqgevfyscepg 288
QY 223 YEYERGDVCTESG-WRP-LPSCSEKSCDN 251
   || || || || || || || || || || || || || || || || || || || || || ||
Db 289 ydlrgaasmrctpgqdwspaaptcevkscdd 319

```

Search completed: November 21, 2000, 16:43:45  
Job time: 284 sec



Tue Nov 21 16:57:27 2000

us-09-316-163-10.rag

Page 11



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:21 ; Search time 27.94 Seconds  
(without alignments)  
376.194 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876  
Sequence: 1 EDCNEIDPPRRNTELTGSGWS.....PDIKHGCLYHNNRRPFV 329

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1876	100.0	1231	1 CFAH_HUMAN	P08603 homo sapien
2	1261	67.2	1234	1 CFAH_MOUSE	P06909 mus musculu
3	353	18.8	597	1 C4BP_HUMAN	P04003 homo sapien
4	350.5	18.7	2039	1 CRL_HUMAN	P17927 homo sapien
5	320.5	17.1	610	1 C4BP_BOVIN	Q28065 bos taurus
6	317	16.9	558	1 C4BP_RAT	Q63514 rattus norv
7	311.5	16.6	469	1 C4BP_MOUSE	P08607 mus musculu
8	307	16.4	1033	1 CR2_HUMAN	P20023 homo sapien
9	303.5	16.2	1025	1 CR2_MOUSE	P19070 mus musculu
10	303	16.2	661	1 F13B_HUMAN	P05160 homo sapien
11	301.5	16.1	263	1 VCP_VACCV	P10998 vaccinia vl
12	294	15.7	668	1 F13B_MOUSE	Q07968 mus musculu
13	286	15.2	830	1 LEM3_HUMAN	P16109 homo sapien
14	279.5	14.9	345	1 APOH_HUMAN	Q01339 mus musculu
15	279	14.9	610	1 LEM2_HUMAN	P16581 homo sapien
16	276.5	14.7	612	1 LEM2_MOUSE	Q00690 mus musculu
17	274.5	14.6	768	1 LEM3_MOUSE	Q01102 mus musculu
18	274	14.6	377	1 MCP_HUMAN	P15529 homo sapien
19	274	14.6	768	1 LEM3_RAT	P98106 rattus norv
20	270.5	14.4	611	1 LEM2_CANFA	P33730 canis fami
21	269.5	14.4	360	1 CCPH_HSVSA	Q01016 herpesvirs
22	267.5	14.3	345	1 APOH_HUMAN	P02749 homo sapien
23	266	14.2	484	1 LEM2_PIG	P98110 sus scrofa
24	262.5	14.0	958	1 HIG_DROME	Q09101 drosophila
25	262	13.9	381	1 DAF_HUMAN	P08174 homo sapien
26	261.5	13.9	345	1 APOH_BOVIN	P17690 bos taurus
27	259	13.8	769	1 LEM3_SHEEP	P98109 ovis aries
28	252	13.4	340	1 DAF_PONPY	P49457 pongo pygma
29	246.5	13.1	507	1 DAF_CANVO	Q60401 cavia porce
30	242.5	12.9	345	1 APOH_CANFA	P33703 canis fami
31	241	12.8	549	1 LEM2_RAT	P98105 rattus norv
32	239.5	12.8	551	1 LEM2_RABIT	P27113 oryctolagus
33	235	12.5	390	1 DAF1_MOUSE	Q61475 mus musculu

## ALIGNMENTS

RESULT 1	ID	CFAH_HUMAN	STANDARD:	PRT: 1231 AA.
AC	P08603	01-AUG-1988 (Rel. 08, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	COMPLEMENT FACTOR H PRECURSOR.			
GN	HFI OR HF OR CFH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LIVER:			
RX	MEDLINE: 86134059.			
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.:			
RT	"The complete amino acid sequence of human complement factor H.;"			
RL	Biochem. J. 249:593-602(1988).			
LN	[2]			
RP	SEQUENCE OF 53-445 FROM N.A.			
RX	MEDLINE: 87054207.			
RA	Schultz T.F., Schaeble W., Stanley K.K., Weiss E., Dierich M.P.:			
RT	"Human complement factor H: Isolation of cDNA clones and partial cDNA sequence of the 38-kDa tryptic fragment containing the binding site for C3b.;"			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
LN	[3]			
RP	SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.			
RX	MEDLINE: 86169701.			
RA	Kristensen T., Wetzel R.A., Tack B.F.:			
RT	"Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.;"			
RL	J. Immunol. 136:3407-3411(1986).			
LN	[4]			
RP	SEQUENCE OF 1047-1231 FROM N.A.			
RX	MEDLINE: 91201892.			
RA	Esteller C., Koistinen V., Schaeble W., Dierich M.P., Weiss E.H.:			
RT	"Cloning of the 1.4-kb mRNA species of human complement factor H reveals a novel member of the short consensus repeat family related to the carboxy terminal of the classical 150-kDa molecule.;"			
RL	J. Immunol. 146:3190-3196(1991).			
LN	[5]			
RP	SEQUENCE OF 19-35.			
RX	MEDLINE: 83048213.			
RA	Sim R.B., Discipio R.G.:			
RT	"Purification and structural studies on the complement-system control protein beta 1H (factor H).;"			
RL	Biochem. J. 205:285-293(1982).			
LN	[6]			
RP	STRUCTURE BY NMR OF 927-985 (SUSHI 16).			
RX	MEDLINE: 91278097.			
RA	Norman D.G., Barlow P.N., Baron M., Day A.J., Campbell I.D.:			
RT	"Three-dimensional structure of a complement control protein module in solution.;"			
RL	J. Mol. Biol. 219:717-725(1991).			

34	233.5	12.4	485	1 LEM2_BOVIN	P98107 bos taurus
35	222.5	11.9	330	1 CPH1_HUMAN	Q03591 homo sapien
36	222.5	11.9	331	1 CPH2_HUMAN	Q02985 homo sapien
37	222.5	11.9	407	1 DAF2_MOUSE	Q61476 mus musculu
38	222	11.8	646	1 LEM3_BOVIN	P42201 bos taurus
39	217	11.6	297	1 APOH_RAT	P26644 rattus norv
40	210	11.2	372	1 LEM1_MOUSE	P18337 mus musculu
41	203.5	10.8	372	1 LEM1_HUMAN	P14151 homo sapien
42	203.5	10.8	372	1 LEM1_PANTR	Q95237 pan troglod
43	202.5	10.8	372	1 LEM1_PONPY	Q95235 pongo pygma
44	200	10.7	372	1 LEM1_RAT	P30836 rattus norv
45	194	10.3	372	1 LEM1_MACMU	Q95198 macaca mula

[7]  
RN STRUCTURE BY NMR OF 264-322 (SUSHI 5).  
RP MEDLINE: 92232649.  
RX Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,  
RA Driscoll P.C., Sim B., Campbell I.D.;  
RT "Solution structure of the fifth repeat of factor H: a second example  
of the complement control protein module.";  
RL Biochemistry 31:3626-3634(1992).  
[8]  
RN STRUCTURE BY NMR OF 866-985 (SUSHIS 15 AND 16).  
RP MEDLINE: 93323119.  
RX Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,  
RA Sim B., Campbell I.D.;  
RT "Solution structure of a pair of complement modules by nuclear  
magnetic resonance.";  
RL J. Mol. Biol. 232:268-284(1993).  
CC -I- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
C3BBB COMPLEX (C3 CONVERTASE) AND THE (C3)NBB COMPLEX (C5  
CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
CC -I- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.  
CC -I- CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION  
341 ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Y00716; CAA68704.1; -;  
DR EMBL: X04697; CAA81739.1; ALT\_FRAME.  
DR EMBL: M65294; AAA35948.1; -;  
DR PIR: S00254; NBRH04.  
DR PIR: S00254; S00254.  
DR PIR: S03013; S03013.  
DR PDB: 1HCC; 15-APR-92.  
DR PDB: 1HFI; 15-JUL-93.  
DR PDB: 1HFI; 15-JUL-93.  
DR MIM: 134370; -;  
DR INTERPRO: IPR000436; -;  
DR PFMW: PFM00084; sush1; 20.  
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;  
Signal; 3D-structure; Polymorphism.  
FT SIGNAL 1 18  
FT CHAIN 19 1231  
FT DOMAIN 20 1229  
FT REPEAT 20 81  
FT REPEAT 84 142  
FT REPEAT 145 206  
FT REPEAT 209 263  
FT REPEAT 324 321  
FT REPEAT 388 386  
FT REPEAT 447 506  
FT REPEAT 508 565  
FT REPEAT 568 624  
FT REPEAT 629 685  
FT REPEAT 690 745  
FT REPEAT 752 804  
FT REPEAT 810 865  
FT REPEAT 869 927  
FT REPEAT 930 985  
FT REPEAT 988 1044  
FT REPEAT 1047 1103  
FT REPEAT 1108 1164  
FT REPEAT 1166 1229  
FT REPEAT 1229 1231  
FT DISULFID 52 80  
FT DISULFID 85 129  
FT DISULFID 114 141

FT DISULFID 146 192 BY SIMILARITY.  
FT DISULFID 178 205 BY SIMILARITY.  
FT DISULFID 210 251 BY SIMILARITY.  
FT DISULFID 237 262 BY SIMILARITY.  
FT DISULFID 267 309 BY SIMILARITY.  
FT DISULFID 294 320 BY SIMILARITY.  
FT DISULFID 325 374 BY SIMILARITY.  
FT DISULFID 357 385 BY SIMILARITY.  
FT DISULFID 389 431 BY SIMILARITY.  
FT DISULFID 416 442 BY SIMILARITY.  
FT DISULFID 448 494 BY SIMILARITY.  
FT DISULFID 477 505 BY SIMILARITY.  
FT DISULFID 509 553 BY SIMILARITY.  
FT DISULFID 536 564 BY SIMILARITY.  
FT DISULFID 569 611 BY SIMILARITY.  
FT DISULFID 597 623 BY SIMILARITY.  
FT DISULFID 630 673 BY SIMILARITY.  
FT DISULFID 659 684 BY SIMILARITY.  
FT DISULFID 691 733 BY SIMILARITY.  
FT DISULFID 719 744 BY SIMILARITY.  
FT DISULFID 753 792 BY SIMILARITY.  
FT DISULFID 781 803 BY SIMILARITY.  
FT DISULFID 811 853 BY SIMILARITY.  
FT DISULFID 839 864 BY SIMILARITY.  
FT DISULFID 870 915 BY SIMILARITY.  
FT DISULFID 901 926 BY SIMILARITY.  
FT DISULFID 931 973 BY SIMILARITY.  
FT DISULFID 959 984 BY SIMILARITY.  
FT DISULFID 989 1032 BY SIMILARITY.  
FT DISULFID 1018 1043 BY SIMILARITY.  
FT DISULFID 1048 1091 BY SIMILARITY.  
FT DISULFID 1077 1102 BY SIMILARITY.  
FT DISULFID 1109 1152 BY SIMILARITY.  
FT DISULFID 1138 1163 BY SIMILARITY.  
FT DISULFID 1167 1218 BY SIMILARITY.  
FT DISULFID 1201 1228 BY SIMILARITY.  
FT CARBOHYD 529 529 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 718 718 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 802 802 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 822 822 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 882 882 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 911 911 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 1029 1029 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT CARBOHYD 1095 1095 N-LINKED (GLCNAc. . .) (POTENTIAL).  
FT VARIANT 402 402 H->Y.  
FT CONFLICT 21 21 /FTID=VAR\_001979.  
FT CONFLICT 30 30 C->Q (IN REF. 3).  
FT CONFLICT 34 34 T->V (IN REF. 3).  
FT CONFLICT 53 54 T->Q (IN REF. 3).  
FT CONFLICT 870 871 RP->IL (IN REF. 2).  
FT STRAND 876 876  
FT TURN 877 878  
FT STRAND 879 881  
FT STRAND 890 891  
FT TURN 893 894  
FT STRAND 896 900  
FT TURN 903 904  
FT STRAND 907 907  
FT STRAND 912 916  
FT TURN 917 918  
FT STRAND 919 920  
FT STRAND 926 926  
SQ SEQUENCE 1231 AA; 139125 MW; C65EC8CF8800B3FD CRC64;

Query Match 100.0%; Score 1876; DB 1; Length 1231;  
Best Local Similarity 100.0%; Pred. No. 6.9e-137;  
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 EDCELPPRRNTEILITGWSODTYPEGTQALYKCRPGYSLGNVIMCRKGEWALNPLR 60  
DB 19 EDCELPPRRNTEILITGWSODTYPEGTQALYKCRPGYSLGNVIMCRKGEWALNPLR 78



```

FT CAROHYD 676 676 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 773 773 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 801 801 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 1030 1030 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 1061 1061 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 1234 AA; 139082 MW; C5AC02F341B957E7 CRC64;

Query Match 67.2%; Score 1261; DB 1; Length 1234;
Best Local Similarity 65.7%; Pred. No. 1.4e-89;
Matches 216; Conservative 37; Mismatches 76; Indels 0; Gaps 0;

OY 1 EDCNELPRRNRNLTGLNSDQTYPPGTOAIYCRPGYSLGNVIMVCKRGWVALNPLR 60
DB 19 EDCKGPPEPNSLISGWSSEQLYRPGTOATYCRPGYRLTGIYVCKKNGKVAASNSR 78
OY 61 KCKRPGHGDPTFTLTGNGVEYGVKAVYTCNNEGOLLGELNRCDDTDGWTNDI 120
DB 79 ICRKRCGHPGDTFPGSFRLAVGSOEFEGAKVYTCDGQLGELIDYRCGADGWTNDI 138
OY 121 PICEVVKCLPYTAENGKIVSSAMEPDRYHFGQAVFYCNSGYKTGDEEMSCDGF 180
DB 139 PICEVVKCLPYTELENGRIYSGAETDQYEGVYRFGCNSFKIEGKHEHCSEGL 198
OY 181 SKEKRCVETISCKSPVINGSPISOKIITKENRFQYKCNMGYSEBQDVCESGWRP 240
DB 199 SNEKRCVETILTTPREVENGDGINVKYKENERHYKCKHGVKPERGAVGTCGSWS 258
OY 241 LPSCEKSCDNPYIPNGDYSPLRIKHTGDEITTCRCNGFVPAFRNTAKDGTIPAP 300
DB 259 GPCEKRCSPYIINGITPHRIIHRSDDELRYECNCFYPTVGTSTVSKCTPTGMI 318
OY 301 RCTLRPCDPTDIKGGLYHENMRPFY 329
DB 319 RCTLRPCDPTDIKGGLYHENMRPFY 347

RESULT 3
C4BP_HUMAN STANDARD; PRT: 597 AA.
AC P04003;
DT 23-OCT-1986 (Rel. 02, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN)
DE (PRT).
GN C4BP OR C4BP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 90073699.
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;
RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT identity of PRP as C4b-binding protein.";
RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 9113199.
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT protein gene.";
RL Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [3]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE: 88242821.
RA Litten S.J., Lewin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT C4b-binding protein and interspecies cross-hybridisation of the C4bp

```

```

RT cDNA sequence.";
RL FEBS Lett. 232:328-332(1988).
RN [4]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE: 86301119.
RA Litten S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RL FEBS Lett. 204:77-81(1986).
RN [5]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE: 86025405.
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RT binding protein, a regulatory protein of the classical pathway of the
RT human complement system.";
RL Biochem. J. 230:133-141(1985).
RN [6]
RP SEQUENCE OF 49-88.
RX MEDLINE: 85296001.
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence and studies of the fragments produced by limited
RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT bromide treatment.";
RL Mol. Immunol. 22:427-435(1985).
RN [7]
RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.
RX MEDLINE: 83221615.
RA Dahlback B., Smith C.A., Mellier-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RT vitamin K-dependent protein S and complement protein C4b.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
CC -I- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOGULANT PROTEIN S
CC AND WITH SERUM AMYLOID P COMPONENT.
CC -I- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1
CC BETA CHAIN, A 530 KDA HOMOPENTAMER OF ALPHA CHAINS OR A 500 KDA
CC COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
CC THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
CC BINDING SITE FOR C4B AT THE END.
CC -I- TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
CC -I- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -I- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M31452; AAA36507.1; -
DR EMBL: M62486; AAA36506.1; -
DR EMBL: M62475; AAA36506.1; JOINED.
DR EMBL: M62476; AAA36506.1; JOINED.
DR EMBL: M62477; AAA36506.1; JOINED.
DR EMBL: M62478; AAA36506.1; JOINED.
DR EMBL: M62479; AAA36506.1; JOINED.
DR EMBL: M62480; AAA36506.1; JOINED.
DR EMBL: M62481; AAA36506.1; JOINED.
DR EMBL: M62482; AAA36506.1; JOINED.
DR EMBL: M62484; AAA36506.1; JOINED.
DR EMBL: M62485; AAA36506.1; JOINED.
DR EMBL: X07853; CA30701.1; -
DR EMBL: X04284; CA30701.1; -
DR EMBL: X04296; CA27839.1; -

```

DR EMBL: X02865; CAA26617.1; -  
 DR PIR: A33568; NBHUC4.  
 DR HSSP: P10998; 1VVC.  
 DR MIM: 120830; -  
 DR INTERPRO: IPR000436; -  
 DR PIR: P00084; sushi: 8.  
 KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;  
 KW Polymorphism.  
 FT SIGNAL 1 48  
 FT CHAIN 49 597 CAB-BINDING PROTEIN ALPHA CHAIN.  
 FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.  
 FT REPEAT 49 109  
 FT REPEAT 112 171 SUSHI 1.  
 FT REPEAT 174 235 SUSHI 2.  
 FT REPEAT 238 295 SUSHI 3.  
 FT REPEAT 298 361 SUSHI 4.  
 FT REPEAT 364 423 SUSHI 5.  
 FT REPEAT 425 481 SUSHI 6.  
 FT REPEAT 483 539 SUSHI 7.  
 FT REPEAT 50 96 SUSHI 8.  
 FT DISULFID 81 108 BY SIMILARITY.  
 FT DISULFID 113 154 BY SIMILARITY.  
 FT DISULFID 140 170 BY SIMILARITY.  
 FT DISULFID 175 217 BY SIMILARITY.  
 FT DISULFID 203 234 BY SIMILARITY.  
 FT DISULFID 239 281 BY SIMILARITY.  
 FT DISULFID 267 294 BY SIMILARITY.  
 FT DISULFID 299 348 BY SIMILARITY.  
 FT DISULFID 332 360 BY SIMILARITY.  
 FT DISULFID 364 7387 BY SIMILARITY.  
 FT DISULFID 7365 7409 BY SIMILARITY.  
 FT DISULFID 7399 422 BY SIMILARITY.  
 FT DISULFID 426 468 BY SIMILARITY.  
 FT DISULFID 454 480 BY SIMILARITY.  
 FT DISULFID 484 525 BY SIMILARITY.  
 FT DISULFID 511 538 BY SIMILARITY.  
 FT DISULFID 546 546 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).  
 FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).  
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).  
 FT VARIANT 92 92 O -> T.  
 FT VARIANT /FTID=VAR\_001977.  
 FT VARIANT Y -> H.  
 FT VARIANT /FTID=VAR\_001978.  
 FT SEQUENCE 597 AA; 67033 MM; 67E03FE2A85A16DD CRC64;  
 SQ  
 Query Match 18.8%; Score 353; DB 1; Length 597;  
 Best Local Similarity 29.38; Pred. No. 4.6e-20;  
 Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;  
 20 SDQYRPEGTQAIYKCRPGY-RSLGNVIMVCRK-GEWALNPLKCKRPGCHGDPTEGT 77  
 67 TETNRKIGTTLTKYCLPEYVSHSTQGLTCSDEWV-YNFE-CIYKCRHPRGE----- 118  
 78 FTLLGANN-----FEYKAVYITNEGQGLGEINRECDTD---GWTNDIPTEVYKC 128  
 119 -LRNGQVEIKTIDLSFGSIEFSCSEGFLLIGSTTSR-CEVQDRGVGSHPLPQCEIVKC 175  
 129 LPVAPRENGKIVSSAMEPDREYHFGQAVRFVNCNGYKIEGDEMCSD---GFMSEK 184  
 176 KPPPIRNGRSHGE-----ENFYAGFSVTYISCDPRFSLGHAISCTYVNETIGVWRPSP 231  
 185 PKVIEISCKSPDVINGSPISQ-KIYKENERFQYKCNMGYSESGDAV--C-TESGWRP 240  
 232 PCKEIKCRKRPDVSHGEWVSGFGPIYKNTIVFCCKGKGFVL--RGSSVICHADADSKMNP 289  
 241 L-PSEEEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITTCGRNGFYAT-RGNMARC- 291  
 290 SPPACEPNSCINLPDIPHASWETYPRTKEDYVVGIVLVRCHPGYKPTTDEPTTIVICQ 349  
 292 TSTGWIPARCTLKPDCYDPIKHGGLYHENMRRP-----YF 327

DR 350 KNLRWTPYGCALCCPEPLNNGEITQHRKSRANHCYVF 390  
 RESULT 4  
 CRL\_HUMAN  
 ID CRL\_HUMAN STANDARD; PRT; 2039 AA.  
 AC P17927;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (C3B/C4B RECEPTOR) (CD35 ANTIGEN).  
 GN CRL OR C3BR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89035992.  
 RA Klickstein L.B., Bartow T.J., Miletic V., Rabson L.D., Smith J.A., Fearon D.T.;  
 RT "Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor (CRL, CD35) by deletion mutagenesis.";  
 RL J. Exp. Med. 168:1699-1717(1988).  
 RN [2]  
 RP SEQUENCE OF 503-2039 FROM N.A.  
 RX MEDLINE: 87168191.  
 RA Klickstein L.B., Wong W.W., Smith J.A., Weis J.H., Wilson J.G., Fearon D.T.;  
 RT "Human C3b/C4b receptor (CRL). Demonstration of long homologous repeating domains that are composed of the short consensus repeats characteristics of C3/C4 binding proteins.";  
 RL J. Exp. Med. 165:1095-1112(1987).  
 RN [3]  
 RP SEQUENCE OF 761-783; 831-845 AND 1179-1195 FROM N.A.  
 RX MEDLINE: 86067975.  
 RA Wong W.W., Klickstein L.B., Smith J.A., Weis J.H., Fearon D.T.;  
 RT "Identification of a partial cDNA clone for the human receptor for complement fragments C3b/C4b.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7711-7715(1985).  
 CC -!- FUNCTION: CRL, PRESENT ON ERYTHROCYTES, LEUCOCYTES, GLOMERULAR PODOCYTES, AND SPLENIC FOLLICULAR DENDRITIC CELLS. MEDIATES THE BINDING BY THESE CELLS OF PARTICLES AND IMMUNE COMPLEXES THAT HAVE ACTIVATED COMPLEMENT.  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- POLYMORPHISM: CRL IS RESPONSIBLE FOR THE KNOX BLOOD GROUP SYSTEM.  
 CC -!- MISCELLANEOUS: SEVEN SHORT CONSENSUS REPEATS (SCR) CONSTITUTE A LONG HOMOLOGOUS REPEAT (LHR). THE N-TERMINAL TWO SRCS OF LHR-A CONTAINED A SITE DETERMINING C4 SPECIFICITY, AND THE N-TERMINAL TWO SRCS OF LHR-B AND -C EACH HAD A SITE DETERMINING C3 SPECIFICITY.  
 CC -!- MISCELLANEOUS: THIS IS THE SEQUENCE OF THE F ALLOTYPE OF CRL.  
 CC -!- SIMILARITY: CONTAINS 30 SUSHI (SCR) REPEATS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M11569; AAA52297.1; -  
 DR EMBL: M11617; AAA52298.1; -  
 DR EMBL: M11618; AAA52299.1; -  
 DR EMBL: Y00816; CAA68755.1; -  
 DR EMBL: X05309; CAA28933.1; -  
 DR PIR: A28507; A28507.  
 DR PIR: A24748; A24748.  
 DR PIR: B24748; B24748.  
 DR PIR: C24748; C24748.

```

DR PIR: S03843; S03843.
DR HSP: P08603; 1HFL.
DR MIM: 120620; -.
DR INTERPRO: IPR000436; -.
DR PFAM: PF00084; sushi; 30.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sushi; Blood group antigen.
FT SIGNAL 1 41
FT CHAIN 42 2039
FT DOMAIN 42 1971
FT TRANSMEM 1972 1996
FT MOD_RES 1997 2039
FT DOMAIN 42 488
FT REPEAT 42 100
FT REPEAT 103 162
FT REPEAT 165 233
FT REPEAT 237 294
FT REPEAT 296 354
FT REPEAT 357 417
FT REPEAT 420 488
FT DOMAIN 492 938
FT REPEAT 492 550
FT REPEAT 553 612
FT REPEAT 615 683
FT REPEAT 687 744
FT REPEAT 746 804
FT REPEAT 807 867
FT REPEAT 870 938
FT DOMAIN 942 1388
FT REPEAT 942 1000
FT REPEAT 1003 1062
FT REPEAT 1065 1133
FT REPEAT 1137 1194
FT REPEAT 1196 1254
FT REPEAT 1257 1317
FT REPEAT 1320 1388
FT DOMAIN 1395 1846
FT REPEAT 1395 1453
FT REPEAT 1456 1515
FT REPEAT 1518 1586
FT REPEAT 1590 1647
FT REPEAT 1649 1707
FT REPEAT 1710 1770
FT REPEAT 1773 1841
FT DOMAIN 1847 1966
FT REPEAT 1847 1905
FT REPEAT 1908 1966
FT DISULFID 43 86
FT DISULFID 73 99
FT DISULFID 104 145
FT DISULFID 131 161
FT DISULFID 166 215
FT DISULFID 195 232
FT DISULFID 238 280
FT DISULFID 266 293
FT DISULFID 297 340
FT DISULFID 326 353
FT DISULFID 358 400
FT DISULFID 386 416
FT DISULFID 421 470
FT DISULFID 450 487
FT DISULFID 493 536
FT DISULFID 523 549
FT DISULFID 554 595
FT DISULFID 581 611
FT DISULFID 616 662
FT DISULFID 645 682
FT DISULFID 686 730

```

```

FT DISULFID 716 743
FT DISULFID 747 790
FT DISULFID 776 803
FT DISULFID 808 850
FT DISULFID 836 866
FT DISULFID 871 920
FT DISULFID 900 937
FT DISULFID 943 966
FT DISULFID 973 999
FT DISULFID 1004 1045
FT DISULFID 1031 1061
FT DISULFID 1066 1115
FT DISULFID 1095 1132
FT DISULFID 1138 1180
FT DISULFID 1166 1193
FT DISULFID 1197 1240
FT DISULFID 1226 1253
FT DISULFID 1258 1300
FT DISULFID 1286 1316
FT DISULFID 1321 1370
FT DISULFID 1350 1387
FT DISULFID 1396 1439
FT DISULFID 1426 1452
FT DISULFID 1457 1498
FT DISULFID 1484 1514
FT DISULFID 1519 1568
FT DISULFID 1548 1585
FT DISULFID 1591 1633
FT DISULFID 1619 1646
FT DISULFID 1650 1693
FT DISULFID 1679 1706
FT DISULFID 1711 1753
FT DISULFID 1739 1769
FT DISULFID 1774 1823
FT DISULFID 1803 1840
FT DISULFID 1848 1891
FT DISULFID 1877 1904
FT DISULFID 1909 1952
FT DISULFID 1938 1965
FT CARBOHYD 56 56
FT CARBOHYD 252 252
FT CARBOHYD 410 410
FT CARBOHYD 447 447
FT CARBOHYD 509 509
FT CARBOHYD 578 578
FT CARBOHYD 702 702
FT CARBOHYD 860 860
FT CARBOHYD 897 897
FT CARBOHYD 959 959
FT CARBOHYD 1028 1028
FT CARBOHYD 1152 1152
FT CARBOHYD 1310 1310
FT CARBOHYD 1481 1481
FT CARBOHYD 1504 1504
FT CARBOHYD 1534 1534
FT CARBOHYD 1540 1540
FT CARBOHYD 1605 1605
FT CARBOHYD 1763 1763
FT CARBOHYD 1908 1908
SQ SEQUENCE 2039 AA; 223589 MW; B2FD29B6AD3C5EB7 CRC64;

Query Match 18.7%; Score 350.5; DB 1; Length 2039;
Best Local Similarity 29.3%; Pred. No. 2.9e-19;
Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE--LPRRNTELTISWSDQTYPECTQAIYKCRPGYRSIGNIVYICRGEWVALNPL 59
   || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 43 CNAPEWLPFAPNTNT---DEFERPICTIYLNBERPGYSGRPSIICLNKSVMTGAAD- 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 60 KCKQKRPQGHGDPPEPGFTLTCGNVFEYGVKAVYTCNEGYQLLGEINVRGC---DTDG 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 -KCRKSKCRNPDPDVGWVHTKG--IOFGSQIKYSCFKGYRLIGS--SSARCIISGDTVI 153

```



Query	Match	Best Local Similarity	Score	DB	Length
Matches	99; Conservative	54; Mismatches	142; Indels	63; Gaps	
QY	6 LPPRNTEPIILGMSDQYPPREGTQAIYKCRGRG--SLGVNIMCRGGEVALNPLTKCQ	63	17.18; 27.78; 54; 142; 63; 610;		
DB	52 IPPYLDEAFPINELNETREFETGLRLTTCRGVYIISKKNFLIDGDGDM--KYKEFCV	108			
QY	64 KRPGHDPDFEFGFTLLTGCNV-----FEYGVKAVVYTCNCEGYDLLEIN-YRECDTG-	115			
DB	109 KKRGNPCE-----LLNGQVIYKTDVSPFSELEFSCSEGYVLIGANSYCOLDQKGVV	161			
QY	116 WTNDIPICEVYVCKLPVTAPENGKVISSAMEDREYHNGQAVRVCNSGKITEGDEHNCS	175			
DB	162 WSDPIPOCIIACEPRPIISNGR--HNGDDE-FTYTGSSVYTCSDDFSLGKASISCR	218			
QY	176 DD---GFWSEKRPCEVCISSKSPVINGSPISOKI-----IKKENRFQYKCNMGXEY	225			
DB	219 VENKTIQVWHSPPCKKAVICYQRPVAKDG-----KITSGEPRITTYQDSIYVACNKGFR	273			
QY	226 SERGDVAV--C--TESGMR-PLPSCERKSC-DNPVYIPNGDYSPLRITKHTGDE-----	271			
DB	274 --EGDSLHCEADNSWNPBPPTCELNGCLGRLPHIALWE--RYDHDQTEQOQVYDIGFV	329			
QY	272 ITYOGRNGEYFATRG--NTAKICST--GMVAPRCITLKPDYDIDIHGG--LVHEMMRRP	325			
DB	330 ISYKCHFGIKPETDEPTVTCOSNLEWSPYTECKEVCCEPFLNNYGSITLH---RRP	384			
RESULT 6					
CABP_RAT	6	STANDARD:	PRT:	558 AA.	
AC	063514:				
DT	01-NOV-1997 (rel. 35, Created)				
DT	01-NOV-1997 (rel. 35, Last sequence update)				
DT	15-JUL-1998 (rel. 36, Last annotation update)				
GN	C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.				
GN	C4BPA.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;				
RC	MEDLINE: 97166082.				
RX	Hillarp A., Wiklund H., Thern A., Dahlback B.;				

RT "Molecular cloning of rat C4b binding protein alpha- and beta-chains: structural and functional relationships among human, bovine, rabbit, mouse, and rat proteins." J. Immunol. 158:1315-1323(1997).

CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3 CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S AND WITH SERUM AMYLOID P COMPONENT.

CC -1- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS.

CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.

CC -1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z50051; CAA90391.1; .

DR HSSP: P10998; IYVC.

DR INTERPRO: IPR000436; .

DR PFAM: PF00084; sushi; 8.

KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.

FT SIGNAL 1 13 BY SIMILARITY.

FT CHAIN 1 558 C4B-BINDING PROTEIN ALPHA CHAIN.

FT DOMAIN 14 502 8 X SUSHI (SCR) REPEATS.

FT REPEAT 14 73 SUSHI 1.

FT REPEAT 76 135 SUSHI 2.

FT REPEAT 138 200 SUSHI 3.

FT REPEAT 203 259 SUSHI 4.

FT REPEAT 262 325 SUSHI 5.

FT REPEAT 328 387 SUSHI 6.

FT REPEAT 389 444 SUSHI 7.

FT REPEAT 446 502 SUSHI 8.

FT DISULFID 15 60 BY SIMILARITY.

FT DISULFID 45 72 BY SIMILARITY.

FT DISULFID 77 118 BY SIMILARITY.

FT DISULFID 104 134 BY SIMILARITY.

FT DISULFID 139 182 BY SIMILARITY.

FT DISULFID 168 199 BY SIMILARITY.

FT DISULFID 204 246 BY SIMILARITY.

FT DISULFID 232 258 BY SIMILARITY.

FT DISULFID 263 312 BY SIMILARITY.

FT DISULFID 296 324 BY SIMILARITY.

FT DISULFID 328 351 BY SIMILARITY.

FT DISULFID 329 373 BY SIMILARITY.

FT DISULFID 363 386 BY SIMILARITY.

FT DISULFID 390 431 BY SIMILARITY.

FT DISULFID 417 443 BY SIMILARITY.

FT DISULFID 447 488 BY SIMILARITY.

FT DISULFID 474 501 BY SIMILARITY.

FT DISULFID 509 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).

FT DISULFID 521 521 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).

FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 558 AA; 62266 MW; 592F0C67ED1E5FF CRC64;

Query Match 16.98; Score 317; DB 1; Length 558;  
Best Local Similarity 27.08; Pred. No. 2.3e-17;  
Matches 110; Conservative 46; Mismatches 128; Indels 124; Gaps 24;

QY 7 PPRNTEILGWSMDOT-YPEGTOAIKCRPGY-RSLGNVIMVCRK-GEW---VALNLR 60  
DB 17 PPDPLPALPASEMNQDDESHITLRINCRCRGYSRASSQSGLYCKPKQWQINTA----- 71

QY 61 KCKRPGHGDTPFGFTLTGQNV-----FEYKAVAVYCNQGYLLG-EINVRCDTD 114  
DB 72 -CKKSGRNGD-----LQNKVEYKIDFLGSGLEFSCSGYILLGSGTSTYCEIQK 123  
QY 115 G--WTDIPICEVYKCLPYTAPENK-----IVSSAMEPD 147  
DB 124 GVSMDPDLCEVIKACGMPDISGNKHNGREEFFTYRSVTKCPDPDFLLGNASTJCT 183  
QY 148 -----REHFQGVAVFYVNSGKIGDEE 171  
DB 184 VVKITGVWSPSPPTCERIICPWPVKVLHGFTNSGFHFKYKDSVRFVQKGFVLGSGV 243  
QY 172 MHCSDDGFWSKEPKCEVISCKS-PRVINGSPIS-----OKIYKENERFOYKCNNGYEY 225  
DB 244 IHCADGSMV-PYVCELSNCTDIPDIPNALITSPRKEDYIPYGVLYRICRGYER 302  
QY 226 SERG--DANCTES-GWRPLPCEKSCDNPYIPNGYSPLR-TKHRT-----GD 270  
DB 303 ATRQPTVLCQKDLWSMRLGCKEICCPVP-----DPKSVRYIQHKAPHDNDCTYFED 357  
QY 271 EITYOCRNGEFYPATRGNTAKCTSTG-WIP-APRCITLPPDDYPD-IKHG 315  
DB 358 EYSYTDNDIM-----LTATCKSDGTWHPRTPSCH-QSCDFPALAHG 399

RESULT 7  
C4BP\_MOUSE STANDARD; PRT; 469 AA.  
ID C4BP\_MOUSE P08607:  
AC 01-AUG-1988 (Rel. 08, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE C4B-BINDING PROTEIN PRECURSOR (C4BP).  
GN C4BPA OR C4BP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 88024997.  
RA Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Tack B.F.;  
RT "cDNA structure of murine C4b-binding protein, a regulatory component  
of the serum complement system.";  
Biochemistry 26:468-467(1987).  
RL Biochemistry 26:468-467(1987).  
CC -1- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT  
ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR  
(C3BINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT  
ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3  
CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA  
CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.  
CC -1- SUBUNIT: HOMOPETAMER; NOT COVALENTLY LINKED. MOUSE LACKS THE  
BETA CHAIN OF C4BP.  
CC -1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.  
CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.  
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M17122; AAA37312.1; ALT\_INTT.  
DR PIR: A27117; NEMSC4.  
DR HSSP: P10998; IYVC.  
DR MGD: MGI:88229; C4BP.  
DR INTERPRO: IPR000436; .  
DR PFAM: PF00084; sushi; 6.  
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal.  
FT SIGNAL 1 56

```

FT CHAIN 57 469 CAB-BINDING PROTEIN.
FT DOMAIN 57 414 6 X SUSHI (SCR) REPEATS.
FT REPEAT 57 116 SUSHI 1.
FT REPEAT 119 177 SUSHI 2.
FT REPEAT 180 241 SUSHI 3.
FT REPEAT 244 300 SUSHI 4.
FT REPEAT 302 356 SUSHI 5.
FT REPEAT 358 413 SUSHI 6.
FT DISULFID 58 103 BY SIMILARITY.
FT DISULFID 88 115 BY SIMILARITY.
FT DISULFID 120 160 BY SIMILARITY.
FT DISULFID 146 176 BY SIMILARITY.
FT DISULFID 181 223 BY SIMILARITY.
FT DISULFID 209 240 BY SIMILARITY.
FT DISULFID 245 287 BY SIMILARITY.
FT DISULFID 273 299 BY SIMILARITY.
FT DISULFID 303 343 BY SIMILARITY.
FT DISULFID 329 355 BY SIMILARITY.
FT DISULFID 359 400 BY SIMILARITY.
FT DISULFID 386 413 BY SIMILARITY.
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 227 227 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 469 AA; 51551 MW; 41E137CB8D8C6321 CRC64;

```

```

Query Match 16.6%; Score 311.5; DB 1; Length 469;
Best Local Similarity 28.5%; Pred. No. 5,5e-17;
Matches 94; Conservative 54; Mismatches 149; Indels 33; Gaps 19;

```

```

QY 7 PPRNRTILLGMSNDQF-YPEGTOAIYKCRPGY-RSLGNVIMVCR-KGEM-VALNPLRKC 62
DB 60 PPAIPALPASDVNKRDESHHTLKTECLPGYRGISRMVYKPSGEWETIS----C 115
QY 63 QKRCFGHPGDTPEGTFTLTGNGVPEYGVKAVYTCNEGYYOLLGEINRECDTDG---WTN 118
DB 116 AKKRCRNPGYLNG---YVNGETITFGSQIEFSCQEGFLLVGS-STSCSEVRGKVAWSN 171
QY 119 DIPCEVYKCLPTAPENKGVSSAMEPDRYHGOAVRFGVNSGKRGEGEEMHCS--- 175
DB 172 PPEPCVIVKCGPPPDISNGK--HSGTDFEYFYNHG--ISYTCDFGFRVLGSPFICGTIVVN 227
QY 176 -DDGFMEKKEKCYEISCKSPDIVNGSPIS-OKITVKEKNEFRFOYKCNMGYSESGDAVC 233
DB 228 KTVYVWSSSPPTCKITISQPNIIHGVIYSGKATYTHRDSVRLACGLNGYLRGRHVEEC 287
QY 234 TESS-WRPLPSCSEKSCD-NFYIPNGDYSPLRIKRRGTDEITYOCRNCFYPATKGNATAK 291
DB 288 QGNGWSSSLPTG-C-DECDLPAIVNGYTVSMVYSKIT--LVITYECDKG-YRLVGAIIISC 343
QY 292 TSTGMT-PAPRCTLKPCDYPRIKHGLYHE 320
DB 344 SFSKWKGTAPCKAL-CQKPEVNGMTLSDE 372

```

```

RESULT 8
CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]

```

```

RP SOURCE FROM N.A.
RX MEDLINE; 89123277.
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RT "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor.";
RL J. Biol. Chem. 264:2118-2125(1989).
RN [2]
RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RX MEDLINE; 86287311.
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology with the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
CC -1- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR
CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR
CC DENDRITIC CELLS OF THE SPLEEN.
CC -1- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA) FAMILY.
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -1- DATABASE: NAME-PROV; NOTE-CD guide CD21 entry;
CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROV/CD/CD21.HTM".
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M26004; AAA35786.1; -
DR EMBL; M26016; AAB04638.1; -
DR EMBL; M24007; AAB04638.1; JOINED.
DR EMBL; M24008; AAB04638.1; JOINED.
DR EMBL; M24009; AAB04638.1; JOINED.
DR EMBL; M24010; AAB04638.1; JOINED.
DR EMBL; M24011; AAB04638.1; JOINED.
DR EMBL; M26009; AAB04638.1; JOINED.
DR EMBL; M26010; AAB04638.1; JOINED.
DR EMBL; M26011; AAB04638.1; JOINED.
DR EMBL; M26012; AAB04638.1; JOINED.
DR EMBL; M26013; AAB04638.1; JOINED.
DR EMBL; M26014; AAB04638.1; JOINED.
DR EMBL; M26015; AAB04638.1; JOINED.
DR PIR; A32036; A32036.
DR PIR; A24319; A24319.
DR PIR; B24319; B24319.
DR PIR; C24319; C24319.
DR PIR; D24319; D24319.
DR PIR; E24319; E24319.
DR PIR; F24319; F24319.
DR HSSP; P10998; IYVC.
DR MIM; 120650; -.
DR INTERPRO; IPR000436; -.
DR PFM; PF00084; sush1; 15.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW Receptor; Sush1.
FT SIGNAL 1 20
FT CHAIN 21 1033 COMPLEMENT RECEPTOR TYPE 2.
FT DOMAIN 21 971 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 972 999 POTENTIAL.
FT DOMAIN 1000 1033 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 971 15 X SUSHI (SCR) REPEATS.
FT REPEAT 22 83 SUSHI 1.
FT REPEAT 90 147 SUSHI 2.
FT REPEAT 153 211 SUSHI 3.
FT REPEAT 214 272 SUSHI 4.

```

```

FT REPEAT 275 343 SUSHI 5.
FT REPEAT 350 407 SUSHI 6.
FT REPEAT 409 467 SUSHI 7.
FT REPEAT 470 523 SUSHI 8.
FT REPEAT 526 594 SUSHI 9.
FT REPEAT 601 658 SUSHI 10.
FT REPEAT 661 715 SUSHI 11.
FT REPEAT 718 780 SUSHI 12.
FT REPEAT 787 844 SUSHI 13.
FT REPEAT 850 908 SUSHI 14.
FT REPEAT 911 969 SUSHI 15.
FT DISULFID 23 65 BY SIMILARITY.
FT DISULFID 51 82 BY SIMILARITY.
FT DISULFID 91 132 BY SIMILARITY.
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 154 197 BY SIMILARITY.
FT DISULFID 183 210 BY SIMILARITY.
FT DISULFID 215 256 BY SIMILARITY.
FT DISULFID 242 271 BY SIMILARITY.
FT DISULFID 276 325 BY SIMILARITY.
FT DISULFID 305 342 BY SIMILARITY.
FT DISULFID 351 393 BY SIMILARITY.
FT DISULFID 379 406 BY SIMILARITY.
FT DISULFID 410 453 BY SIMILARITY.
FT DISULFID 439 466 BY SIMILARITY.
FT DISULFID 471 509 BY SIMILARITY.
FT DISULFID 495 522 BY SIMILARITY.
FT DISULFID 527 576 BY SIMILARITY.
FT DISULFID 556 593 BY SIMILARITY.
FT DISULFID 602 644 BY SIMILARITY.
FT DISULFID 630 657 BY SIMILARITY.
FT DISULFID 662 699 BY SIMILARITY.
FT DISULFID 685 714 BY SIMILARITY.
FT DISULFID 719 762 BY SIMILARITY.
FT DISULFID 748 779 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 851 894 BY SIMILARITY.
FT DISULFID 880 907 BY SIMILARITY.
FT DISULFID 912 955 BY SIMILARITY.
FT DISULFID 941 968 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 667 667 O -> D (IN REF. 2).
FT CONFLICT 902 902 O -> G (IN REF. 2).
FT CONFLICT 906 906 H -> L (IN REF. 2).
SQ SEQUENCE 1033 AA; 112973 MW; 1749DBA07847ADA CRC64;

```

Query Match 16.4%; Score 307; DB 1; Length 1033;  
 Best Local Similarity 23.7%; Pred. No. 3e-16;  
 Matches 97; Conservative 58; Mismatches 139; Indels 116; Gaps 22;

```

QY 1 EDNCNELPRRNTLITLGSMDQ---TYPEGQAIKCRPGYSLGNVAVC-RKGEWAL 56
DQ 408 KEC-QAP----NILNGQKEDRHVRPDGTSIKSCNPGLVVEESIQCTSEGVWP- 461
QY 57 NPLAKCKORPGHGS---DIPFGFTLITGNGVFEYGVKAVYTCNEGYQLLGEINVRCD 112
DQ 462 -PVGQCKVAACEATGROLITRPOHOFVRPDV-----SSCGEGYKLSGSV-YQECQ 510
QY 113 -TDGWTNDIPICEVVKCLPVTAPENGKIVSSAMPEPDREYHFGQAVRVCNSG-----YK 165
DQ 511 GTIMFMEIRICKETTCPPVYITNGAHTGSSLE---DPIGTITVYTCNCPGERGVEFS 567

```

```

QY 166 IEGDEMHCSDD---GFWSEKPKC---VEISCKSPDVINGSPISOK-IITYENERFO 216
DQ 568 LIGESTIRCTNSOERBTGWSGAPLCKLSLAVCCSHVHLANGIKISGKEAPFYNDJVT 627
QY 217 YKCMNGYEYSERGDVAVC-TESGWRP-LPSCSEKSDN----- 251
DQ 628 FKCYSGFTLKGSSQIRCKADNTMDPEIPVCEKEFCQHVRSLOELPAGSRVELVNTSCOD 687
QY 252 -----PIIPMGDISPLRIKA-RIGDEITYQ 275
DQ 688 GYOLTGAIYQMODAENGIFPKIPLCKVIHCPPVIVNGKHTGMAENFLYGNEVSYE 747
QY 276 CRNGEYPATRGNTAKCTST-----GWI-PAPRCLTKP-----CDYDRIKHG 315
DQ 748 CDQGEY-LIGEKKLQCRSDSKSGHSGSPDCLRSPPVRCNPVEYKHG 796

RESULT 9
ID CR2_MOUSE STANDARD; PRT; 1025 AA.
AC P19070;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).
GN CR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA MEDLINE; 90229735.
RA Fingerboth J.D.;
RT "Comparative structure and evolution of murine CR2. The homolog of
  the human C3d/EBV receptor (CD21).";
RL J. Immunol. 144:3458-3467(1990).
RN [2]
RP SEQUENCE OF 12-1025 FROM N.A.
RA MEDLINE; 91010789.
RA Molina H., Kinoshita T., Inoue K., Garel J.C., Holers V.M.;
RT "A molecular and immunochemical characterization of mouse CR2.
  Evidence for a single gene model of mouse complement receptors 1 and
  2.";
RL J. Immunol. 145:2974-2983(1990).
RN [3]
RP SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RA MEDLINE; 89098890.
RA Fingerboth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT "Identification of murine complement receptor type 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN [4]
RP SEQUENCE OF 289-1025 FROM N.A.
RA MEDLINE; 89381350.
RA Kurtz C.B., Paul M.S., Aegeerter M., Weis J.J., Weis J.H.;
RT "Murine complement receptor gene family. II. Identification and
  characterization of the murine homolog (Cr2) to human CR2 and its
  molecular linkage to Crry.";
RL J. Immunol. 143:2058-2067(1989).
RN [5]
RP FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
  LYMPHOCYTES ACTIVATION.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC -1- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC -1- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M81083; AAA37451.1; -			
DR	EMBL; M35684; AAA37448.1; -			
DR	EMBL; M61132; AAA63293.1; -			
DR	EMBL; M35685; AAA37450.1; ALT_SEQ.			
DR	EMBL; M29281; AAA37447.1; -			
DR	PIR; A43526; A43526.			
DR	HSSP; P10998; 1VVC.			
DR	MED; MGI:88489; CR2.			
DR	INTERPRO; IPR000436; -			
DR	PFAM; PF00084; sushi; 14.			
KW	Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal; Receptor; Sushi.			
KV	SIGNAL	1	11	
FT	CHAIN	12	1025	COMPLEMENT RECEPTOR TYPE 2.
FT	DOMAIN	12	963	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	964	990	POTENTIAL.
FT	DOMAIN	991	1025	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	13	959	15 X SUSHI (SCR) REPEATS.
FT	REPEAT	13	74	SUSHI 1.
FT	REPEAT	81	139	SUSHI 2.
FT	REPEAT	145	203	SUSHI 3.
FT	REPEAT	206	264	SUSHI 4.
FT	REPEAT	267	335	SUSHI 5.
FT	REPEAT	342	399	SUSHI 6.
FT	REPEAT	401	458	SUSHI 7.
FT	REPEAT	462	515	SUSHI 8.
FT	REPEAT	518	586	SUSHI 9.
FT	REPEAT	593	649	SUSHI 10.
FT	REPEAT	653	705	SUSHI 11.
FT	REPEAT	708	769	SUSHI 12.
FT	REPEAT	777	834	SUSHI 13.
FT	REPEAT	840	898	SUSHI 14.
FT	REPEAT	901	959	SUSHI 15.
FT	DISULFID	14	56	BY SIMILARITY.
FT	DISULFID	42	73	BY SIMILARITY.
FT	DISULFID	82	124	BY SIMILARITY.
FT	DISULFID	110	138	BY SIMILARITY.
FT	DISULFID	146	189	BY SIMILARITY.
FT	DISULFID	175	202	BY SIMILARITY.
FT	DISULFID	207	248	BY SIMILARITY.
FT	DISULFID	234	263	BY SIMILARITY.
FT	DISULFID	268	317	BY SIMILARITY.
FT	DISULFID	297	334	BY SIMILARITY.
FT	DISULFID	343	385	BY SIMILARITY.
FT	DISULFID	371	398	BY SIMILARITY.
FT	DISULFID	402	445	BY SIMILARITY.
FT	DISULFID	431	458	BY SIMILARITY.
FT	DISULFID	463	501	BY SIMILARITY.
FT	DISULFID	487	514	BY SIMILARITY.
FT	DISULFID	519	568	BY SIMILARITY.
FT	DISULFID	548	585	BY SIMILARITY.
FT	DISULFID	594	636	BY SIMILARITY.
FT	DISULFID	622	649	BY SIMILARITY.
FT	DISULFID	654	689	BY SIMILARITY.
FT	DISULFID	709	752	BY SIMILARITY.
FT	DISULFID	738	769	BY SIMILARITY.
FT	DISULFID	778	820	BY SIMILARITY.
FT	DISULFID	806	833	BY SIMILARITY.
FT	DISULFID	841	884	BY SIMILARITY.
FT	DISULFID	870	897	BY SIMILARITY.
FT	DISULFID	902	945	BY SIMILARITY.
FT	DISULFID	931	958	BY SIMILARITY.
FT	CARBOHYD	77	77	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	113	113	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	216	276	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	364	364	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	380	380	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	527	527	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match	16.2%	Score 303.5	DB 1	Length 1025
Best Local Similarity	27.7%	Pred. No. 5,5e-16		
Matches 96	Conservative 37	Mismatches 146	Indels 67	Gaps 22
QY 2	DCNELPRRNTIELTIGSWSDQTYPEGTOAIYKCRQYRSLGNVIYVC-RKGEVALNPLR 60			
DB 145	FCPSLPTLNHGH-HTGHQVWDQ-FVAGLSLVTYCECEPGYLLTGKTKTICLSSGMDGVIIP- 200			
QY 61	KQKRPCGHPGDTPTGCTLTLLGNN-----FEYGVKAVYTCNEGVLGELIYREC----- 111			
DB 201	TCKEACQCEHPGKFP-----NGQYKEPISLDVGTTFVSCNEGLOLQSQPS-SQCVIYE 252			
QY 112	DDVDGNDIPICEFYVYKCLVTPABENCKIYSSAMEPDRREYHFGQAVRFQNS-----GYK 165			
DB 253	QKAIWTKK-PYCKEILCLPPRPYRANSHTGSPSE---NVPYGSVTIYTCDPSEKGVST 308			
QY 166	IRDEEMHCS---DDGFWSKERPKCVY---ETSCKSPYVINGSPIS-QKIYKENERO 216			
DB 309	LIGKEKINTCTGSQKTGYMSGAPYCVLSTSAVLCLQPKIKRQGLILKSDSYNDTVA 368			
QY 217	YKCNMGYKESRDQAVCTSSG-WR-PLPSCSEKSCDNP-----YIPNGDYSPLR 263			
DB 369	FCSEPEFTLAKNRSISCNAGHTEWPPVPC-EKGCQAPPKIINGQKEDSYLINF----- 422			
QY 264	IKHRTDEITYOCRNGFYPATRGNTAKCTSTG-WID-APRCTLKPC 307			
DB 423	---PQTSIRYSCDPG-YLLGVEDITHCTPEGKWPITPTQCTVAEC 463			
RESULT 10				
F13B_HUMAN				
ID F13B_HUMAN	STANDARD:	PRT:	661 AA.	
AC P05160:				
DT 13-AUG-1987 (Rel. 05, Created)				
DT 01-AUG-1990 (Rel. 15, Last sequence update)				
DT 15-JUL-1999 (Rel. 38, Last annotation update)				
DE COAGULATION FACTOR XIIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-GLUTAMINASE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B CHAIN).				
DE F13B.				
GN F13B.				
OS Homo sapiens (Human).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
RN (1)				
RP SEQUENCE FROM N.A.				
RX MEDLINE; 91105054.				
RA Botenous R.E., Ichinose A., Davie E.W.;				
RT "Nucleotide sequence of the gene for the b subunit of human factor XIIII.";				
RL Biochemistry 29:11195-11209(1990).				
RN (2)				
RP SEQUENCE OF 2-661 FROM N.A.				
RX MEDLINE; 87026535.				
RA Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;				
RT "Amino acid sequence of the b subunit of human factor XIIII, a protein composed of ten repetitive segments.";				
RL Biochemistry 25:4633-4638(1986).				

FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	654	694	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	813	813	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	901	901	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	289	291	YGS -> EFR (IN REF. 4).	
FT	CONFLICT	306	306	S -> T (IN REF. 2).	
FT	CONFLICT	520	520	P -> A (IN REF. 2).	
FT	CONFLICT	962	963	MISSING (IN REF. 4).	
SO	SEQUENCE	1025 AA;	112994 MW;	19E518B9A0273694 CRC64;	

Query Match	16.2%;	Score 303.5;	DB 1;	Length 1025;
Best Local Similarity	27.7%;	Pred. No. 5.5e-16;		
Matches 96;	Conservative 37;	Mismatches 146;	Indels 67;	Gaps 22;

QY	2	DCNELPRRNTIELTIGSWSDQTYPEGTQAIYKCRQYRSLGNVIMYC-RKGEVALLNPLR	60
DB	145	ECPSLPTLNHGH-HTGHQHWDO-FVAGLSLVTYCECEPGYLLTGKTKTICLSSGMDGVI	200
QY	61	KQKRPCGHPGDTPTGCTLTLLGNN-----FEYGVKAVYTCNEGVLGELIYREC----	111
DB	201	TCKEAOCEHPGRF-----NGQKKEPISLDVGTTVFSCNEGLOLQSQPS-SQCVIYE	252
QY	112	DDWDGNDIPICEFYVYKCLVTPABENCKIYSSAMEPDRREYHFGQAVRFQNS-----GYK	165
DB	253	QKAIMTKK-PYCKEILCLPPRPYRANSHTGSPSE---NVPYSTVATYTCDDPSKECVST	308
QY	166	IBDEEMHCS---DDGFWSKERPKCV-----ETSCKSPYVINGSPS-QKIYKENERO	216
DB	309	LIGKEKINTCTGSQKTGIMSGAPYCVLSTSAVLCLQPKIKRQGLILKIDSYNDTVA	368
QY	217	YKCNMGYKESRDQAVCTSSG-WR-PLPSCSEKSCDNP-----YIPNGDYSPLR	263
DB	369	FSCEPFTLKGNRSICNAHGTEWPPVPC-EKGCQAPPKIINGQKEDSYLLNFD-----	422
QY	264	IKHRTDEITYOCRNGFYPATRGNTAKCTSTG-WID-APRCTLKPC	307
DB	423	---PQTSIRYSCDPG-YLLGVEDTHTCTPEGKWPITPTQCTVAEC	463

RESULT 10			
FL3B_HUMAN	STANDARD;	PRT;	661 AA.
ID	FL3B_HUMAN		
AC	P05160;		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	COAGULATION FACTOR XIIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-		
DE	GLUTAMININE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B		
DE	CHAIN).		
GN	FL3B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RN	(1)		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE; 91105054.		
RA	Botenous R.E., Ichinose A., Davie E.W.;		
RT	"Nucleotide sequence of the gene for the b subunit of human factor		
RT	XIIII.":		
RL	Biochemistry 29:11195-11209(1990).		
RN	(2)		
RN	SEQUENCE OF 2-661 FROM N.A.		
RP	MEDLINE; 87026535.		
RA	Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;		
RT	"Amino acid sequence of the b subunit of human factor XIIII, a protein		
RT	composed of ten repetitive segments.":		
RL	Biochemistry 25:4633-4638(1986).		

FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	654	694	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	813	813	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	901	901	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	289	291	YGS -> EFR (IN REF. 4).	
FT	CONFLICT	306	306	S -> T (IN REF. 2).	
FT	CONFLICT	520	520	P -> A (IN REF. 2).	
FT	CONFLICT	962	963	MISSING (IN REF. 4).	
SO	SEQUENCE	1025 AA;	112994 MW;	19E518B9A0273694 CRC64;	

Query Match	16.2%;	Score 303.5;	DB 1;	Length 1025;
Best Local Similarity	27.7%;	Pred. No. 5.5e-16;		
Matches 96;	Conservative 37;	Mismatches 146;	Indels 67;	Gaps 22;

QY	2	DCNELPRRNTIELTIGSWSDQTYPEGTQAIYKCRQYRSLGNVIMYC-RKGEVALLNPLR	60
DB	145	ECPSLPTLNHGH-HTGQHWDO-FVAGLSVATYCECEPGYLLTGKRTIKLSSGMDGVI	200
QY	61	KQKRPCGHPGDTPTGCTFLTLGGNV-----FEYGVKAVYTCNEGVLGELIYREC----	111
DB	201	TCKEAOCEHPGRF-----NGQKKEPISLDVGTTFVSCNEGLOLQSQPS-SQCVIYE	252
QY	112	DDWDGNDIPICEFYVYKCLPTAPBENCKIYSSAMEPDRREYHFGQAVRFQNS-----GYK	165
DB	253	QKAIMTKK-PYCKEILCLPPRPYRANSHTGSPSE---NVPYSTVATYTCDDPSKECVST	308
QY	166	IBDEEMHCS---DDGFWSKERPKCV-----ETSCKSPVYINGSPIS-QKIYKENERO	216
DB	309	LIGKEKINTCTGSQKTYGIMSGAPYCVLSTSAVLCLQPKIKRQGLILKLDYSYNDTVA	368
QY	217	YKCNMGYKESRDQAVCTSSG-WR-PLPSCSEKSCDNP-----YIPNGDYSPLR	263
DB	369	FSCEPFTLKGNRSICNAHGHTPEVPVPC-EKGCQAPPKIINGQKEDSYLINF-----	422
QY	264	IKHRTDEITYOCRNGFYPATRGNTAKCTSTG-WID-APRCTLKPC	307
DB	423	---PQTSIRYSCDPG-YLLGVEDLTHCTPEGKWPITPTQCTVAEC	463

RESULT 10			
FL3B_HUMAN	STANDARD;	PRT;	661 AA.
ID	FL3B_HUMAN		
AC	P05160;		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	COAGULATION FACTOR XIIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-		
DE	GLUTAMININE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B		
DE	CHAIN).		
GN	FL3B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RN	(1)		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE; 91105054.		
RA	Botenous R.E., Ichinose A., Davie E.W.;		
RT	"Nucleotide sequence of the gene for the b subunit of human factor		
RT	XIIII.;"		
RL	Biochemistry 29:11195-11209(1990).		
RN	(2)		
RN	SEQUENCE OF 2-661 FROM N.A.		
RP	MEDLINE; 87026535.		
RA	Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;		
RT	"Amino acid sequence of the b subunit of human factor XIIII, a protein		
RT	composed of ten repetitive segments.;"		
RL	Biochemistry 25:4633-4638(1986).		

FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	639	639	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	654	694	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	754	754	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	813	813	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	823	823	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	851	851	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	901	901	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	289	291	YGS -> EFR (IN REF. 4).	
FT	CONFLICT	306	306	S -> T (IN REF. 2).	
FT	CONFLICT	520	520	P -> A (IN REF. 2).	
FT	CONFLICT	962	963	MISSING (IN REF. 4).	
SO	SEQUENCE	1025 AA;	112994 MW;	19E518B9A0273694 CRC64;	

Query Match	16.2%;	Score 303.5;	DB 1;	Length 1025;
Best Local Similarity	27.7%;	Pred. No. 5.5e-16;		
Matches 96;	Conservative 37;	Mismatches 146;	Indels 67;	Gaps 22;

QY	2	DCNELPRRNTIELTIGSWSDQTYPEGTQAIYKCRQYRSLGNVIMYC-RKGEVALLNPLR	60
DB	145	ECPSLPTLNHGH-HTGHQHWDO-FVAGLSLVTYCECEPGYLLTGKTKTICLSSGMDGVI	200
QY	61	KQKRPCGHPGDTPTGCTLTLLGNN-----FEYGVKAVYTCNEGVLGELIYREC----	111
DB	201	TCKEAOCEHPGRF-----NGQKKEPISLDVGTTVFSCNEGLOLQSQPS-SQCVIYE	252
QY	112	DDWDGNDIDPICEFYVYKCLVTPABENCKIYSSAMEPDRREYHFGQAVRFQNS-----GYK	165
DB	253	QKAIWTKK-PYCKEILCLPPRPYRANSHTGSPSE---NVPYSTVATYTCDDPSKECVST	308
QY	166	IBDEEMHCS---DDGFWSKERPKCV-----ETSCKSPYVINGSPS-QKIYKENERO	216
DB	309	LIGKEKINTCTGSQKTGIMSGAPYCVLSTSAVLCLQPKIKRQGLILKIDSYNDTVA	368
QY	217	YKCNMGYKXSERODAVCTSSG-WR-PLPSCSEKSCDNP-----YIPNGDYSPLR	263
DB	369	FSCEPFTLKGNRSICNAHGTMEVPVPC-EKGCQAPPKIINGQKEDSYLLNFD-----	422
QY	264	IKHRTDEITYOCRNGFYPATRGNTAKCTSTG-WID-APRCTLKPC	307
DB	423	---PQTSIRYSCDPG-YLLGVEDLTHCTPBEKWPITPTQCTVAEC	463

RESULT 10			
FL3B_HUMAN	STANDARD;	PRT;	661 AA.
ID	FL3B_HUMAN		
AC	P05160;		
DT	13-AUG-1987 (Rel. 05, Created)		
DT	01-AUG-1990 (Rel. 15, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	COAGULATION FACTOR XIIII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-		
DE	GLUTAMININE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGLUTAMINASE B		
DE	CHAIN).		
GN	FL3B.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
RN	(1)		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE; 91105054.		
RA	Botenous R.E., Ichinose A., Davie E.W.;		
RT	"Nucleotide sequence of the gene for the b subunit of human factor		
RT	XIIII.":		
RL	Biochemistry 29:11195-11209(1990).		
RN	(2)		
RN	SEQUENCE OF 2-661 FROM N.A.		
RP	MEDLINE; 87026535.		
RA	Ichinose A., McMullen B.A., Fujikawa K., Davie E.W.;		
RT	"Amino acid sequence of the b subunit of human factor XIIII, a protein		
RT	composed of ten repetitive segments.":		
RL	Biochemistry 25:4633-4638(1986).		

FT	DISULFID	454	505	BY SIMILARITY.
FT	DISULFID	456	515	BY SIMILARITY.
FT	DISULFID	524	567	BY SIMILARITY.
FT	DISULFID	524	578	BY SIMILARITY.
FT	DISULFID	582	636	BY SIMILARITY.
FT	DISULFID	162	646	BY SIMILARITY.
FT	CARBOHYD	162	162	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	545	545	N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	SITE	617	619	CELL ATTACHMENT SITE.
FT	VARIANT	450	450	C-> F (IN F13 DEFICIENCY).
FT	SEQUENCE	661 AA:	75491 MW;	/FTID=VAR_007475.
SO	SEQUENCE	661 AA:	75491 MW;	57A2FB46560857F2 CRC64;
Query Match	Best Local Similarity	16.2%;	Score 303;	DB 1; Length 661;
Matches	93; Conservative	49;	Mismatches 136;	Indels 74; Gaps 20.
OY	13 ELITGWS-	---DOTYPEGTOAIKCPGYKSLNVMVCKRGWV-----	ALNPLRKCQ 63	
Db	157 ELYNGN	STYQTFKVKVQYECAMGYTTAGG-----	KTEDEVECLTYGWSLTP--KCT 209	
OY	64 KRPGC-	-----HFGDTPRGTTLTGNAVFEYGVKAVYTCNCGYLLGEINRBCDT 113		
Db	210 KIKCSRLRIENGYPHPK-	-----QYEEGDVQVQFCHENYULSGS-DLIQCYN 257		
OY	114 DGMTDIPICE-	---VYKCLPTVAPENGKIIVSSAMEDREHFGQAVRFVNSGYKIEGDEE 171		
Db	258 FGWYESPVECBERRRRCRPPRLPIKSIQTHS----	TYRHGEIYHICELNFEIHGSAE 313		
OY	172 MHCSDDGFWSKEPKVC-	---ISCKSPDVI--NGSPISOKIYKENERPOYKCNNGYEYS 226		
Db	314 IRC-EDGKWT-	EPKPCIEGQEKVACEEPPFIENGAAHLHSKIYYNGDKYTAACKSGYLLH 371		
OY	227 ERGADVCHESGRPLPSCDEE--	KSCDN--PIYIPNGYSP--LRKHHTGSEITITQCANGYTP 282		
Db	372 GSNELTTCRACKWTLTP	BECEVENNENKHPYVNGAVADGILASVATGSSVEYRC--NEYLL 430		
OY	283 AIRGNTAKCTSTGWIAPARCTLKPC----	DYPD-----IRHGL 317		
Db	431 LRGSKISRCEGQKMWSPVC-	LEPCTVAVDVMNRNINIMKKMYECKVLHGDL 481		
RESULT 11	VCP_VACCV	STANDARD:	PRT:	263 AA.
AC	P10998;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)			
GN	C3L.			
OS	Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).			
CC	Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	orthopoxvirus.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.			
RC	STRAIN-WR:			
RX	MEDLINE: 88318974.			
RA	Kotwal G.J., Moss B.;			
RT	"Vaccinia virus encodes a secretory polypeptide structurally related			
RL	to complement control proteins.";			
RL	Nature 335:176-178(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-WR:			
RX	MEDLINE: 89073756.			
RA	Kotwal G.J., Moss B.;			
RT	"Analysis of a large cluster of nonessential genes deleted from a			
RL	vaccinia virus terminal transposition mutant.";			
RL	Virology 167:524-537(1988).			
RN	[3]			

RP SEQUENCE FROM N.A.  
 RC STRAIN-COPENHAGEN;  
 RA MEDLINE; 91021027.  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RT "The complete DNA sequence of vaccinia virus.";  
 RL Virology 179:247-266(1990).  
 RL [4]  
 RP COMPLETE GENOME.  
 RC STRAIN-COPENHAGEN;  
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
 RA Paolletti E.;  
 RL Virology 179:517-563(1990).  
 RL [5]  
 RP FUNCTION.  
 RA MEDLINE; 92115714.  
 RA Isaacs S.N., Kotwal G.J., Moss B.;  
 RT "Vaccinia virus complement-control protein prevents  
 RT antibody-dependent complement-enhanced neutralization of infectivity  
 RT and contributes to virulence.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).  
 RN [6]  
 RN STRUCTURE BY NMR OF 146-263.  
 RX MEDLINE; 97446168.  
 RA Wiles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,  
 RA Barlow P.N.;  
 RT "NMR studies of a viral protein that mimics the regulators of  
 RT complement activation.";  
 RL J. Mol. Biol. 272:253-265(1997).  
 CC -1- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY  
 CC INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT  
 CC ACTIVATION. BINDS C3B AND C4B.  
 CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
 CC COMPLEMENT ACTIVATION (RCA).  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13166; CAA31564.1; -;  
 DR EMBL; M2812; AAA69605.1; -;  
 DR EMBL; M35027; AAA47997.1; -;  
 DR PIR; A31005; WMVZSP.  
 DR PDB; 1VVC; 03-DEC-97.  
 DR PDB; 1VVD; 03-DEC-97.  
 DR PDB; 1VVE; 03-DEC-97.  
 DR INTERPRO; IPR000436; -;  
 DR PFAM; PF00084; Sush1; 4.  
 DR Signal; Repeat; Sush1; 3D-structure.  
 FT SIGNAL 1 19  
 FT CHAIN 20 263 COMPLEMENT CONTROL PROTEIN.  
 FT DOMAIN 20 262 4 X SUSHI (SCR) REPEATS.  
 FT REPEAT 20 82 SUSHI 1.  
 FT REPEAT 85 144 SUSHI 2.  
 FT REPEAT 147 202 SUSHI 3.  
 FT REPEAT 205 262 SUSHI 4.  
 FT DISULFID 21 70 BY SIMILARITY.  
 FT DISULFID 54 81 BY SIMILARITY.  
 FT DISULFID 86 126 BY SIMILARITY.  
 FT DISULFID 112 143 BY SIMILARITY.  
 FT DISULFID 148 190 BY SIMILARITY.  
 FT DISULFID 176 201 BY SIMILARITY.  
 FT DISULFID 206 248 BY SIMILARITY.  
 FT DISULFID 234 261 BY SIMILARITY.  
 FT DISULFID 261 261 BY SIMILARITY.  
 SQ SEQUENCE 263 AA; 28629 MW; EA322CC9A6EF997 CRC64;

Query Match

16.1%; Score 301.5; DB 1; Length 263;

Best Local Similarity 33.3%; Pred. No. 1.7e-16;  
 Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;  
 Oy 20 SDQTPPEGTQAIKCRPGYR--SLGNVIMCRKKEGVALNPLRKCKORPCGHPGDTFCT 77  
 Db 40 ANANNIGDTIEYLCLPGRKOKMGPYAKCTGTWTLFN---QC1KRCRSPRIDNQ 96  
 Oy 78 FTLTGNNPFEYGVKAVYCNNEGQILGEL-NYRECDTDS---WTNDIPICEVVKCLPYTA 133  
 Db 97 LDIGG---VDFGSSITYSNCSGYHLGESSKSYCELGSTGSMWNPAPICESVKQSPPS 153  
 Oy 134 PENGKIVSSAMPDEHYHFGQAVRFVNCNGYKIEGDEHMCSDDGFSKKEPKVEISCK 193  
 Db 154 ISNGR--HNGYED--FYTGSGVYITSCNNGYGLIGNSGVLC--GGKMS--DPTCQIVKCP 207  
 Oy 194 SPDVINGSPISQ-KIITYKENERFYKCNNGYEXSERGDVCTESG-WRP-LDSC 244  
 Db 208 HPTISNGYLSGFKRSYSYNDVDFCKYKGLSGSSSTCSFGNTWKPELKC 261  
 RESULT 12  
 ID F13B\_MOUSE STANDARD; PRT; 668 AA.  
 AC 007968;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE COAGULATION FACTOR XIII B CHAIN PRECURSOR (BC 2.3.2.13) (PROTEIN-  
 DE GLUTAMINE GAMMA-GLUTAMYLTRANSFERASE B CHAIN) (TRANSGUTAMINASE B  
 DE CHAIN).  
 GN F13B OR F13B.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN-B10.D2/OSN; TISSUE=LIVER;  
 RX MEDLINE; 93224141.  
 RA Nonaka M., Matsuda Y., Shirosaki T., Moriwaki K., Nonaka M.,  
 RA Natsunne-Sakai S.;  
 RT "Molecular cloning of the b subunit of mouse coagulation factor XIII  
 RT and assignment of the gene to chromosome 1: close evolutionary  
 RT relationship to complement factor H.";  
 RL Genomics 15:535-542(1993).  
 CC -1- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,  
 CC BUT IS THOUGHT TO STABILIZE THE A SUBUNIT AND REGULATE THE RATE  
 CC OF TRANSGUTAMINASE FORMATION BY THROMBIN.  
 CC -1- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.  
 CC -1- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D10071; BAA00963.1; -;  
 DR PIR; A46013; A46013.  
 DR HSSP; P10998; 1VVC.  
 DR MGD; MGI:88379; F13B.  
 DR INTERPRO; IPR000436; -;  
 DR PFAM; PF00084; Sush1; 8.  
 DR Transferrase; Plasma; blood coagulation; Repeat; Glycoprotein; Signal;  
 KW Sush1.  
 FT SIGNAL 1 20  
 FT CHAIN 21 668 BY SIMILARITY.  
 FT DOMAIN 24 647 COAGULATION FACTOR XIII B CHAIN.  
 FT REPEAT 24 88 SUSHI 1.  
 FT REPEAT 90 147 SUSHI 2.  
 FT REPEAT 152 209 SUSHI 3.



```

FT REPEAT 212 268 SUSHI 4.
FT REPEAT 273 328 SUSHI 5.
FT REPEAT 335 390 SUSHI 6.
FT REPEAT 395 451 SUSHI 7.
FT REPEAT 453 516 SUSHI 8.
FT REPEAT 523 579 SUSHI 9.
FT REPEAT 581 647 SUSHI 10.
FT DISULFID 225 76 BY SIMILARITY.
FT DISULFID 39 87 BY SIMILARITY.
FT DISULFID 91 135 BY SIMILARITY.
FT DISULFID 118 146 BY SIMILARITY.
FT DISULFID 153 197 BY SIMILARITY.
FT DISULFID 180 208 BY SIMILARITY.
FT DISULFID 213 255 BY SIMILARITY.
FT DISULFID 241 267 BY SIMILARITY.
FT DISULFID 274 316 BY SIMILARITY.
FT DISULFID 302 327 BY SIMILARITY.
FT DISULFID 336 378 BY SIMILARITY.
FT DISULFID 364 389 BY SIMILARITY.
FT DISULFID 396 439 BY SIMILARITY.
FT DISULFID 425 450 BY SIMILARITY.
FT DISULFID 454 505 BY SIMILARITY.
FT DISULFID 486 515 BY SIMILARITY.
FT DISULFID 524 567 BY SIMILARITY.
FT DISULFID 553 578 BY SIMILARITY.
FT DISULFID 582 636 BY SIMILARITY.
FT DISULFID 616 646 BY SIMILARITY.
FT CARBOHYD 162 162 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 668 AA; 76078 MW; 80BC9E00A9E53FA6 CRC64;

```

```

Query Match 15.7%; Score 294; DB 1; Length 668;
Best Local Similarity 28.1%; Pred. No. 1,8e-15;
Matches 93; Conservative 41; Mismatches 141; Indels 56; Gaps 21;

```

```

OY 13 ELITGMS--DQTPRGTOAIYKCRPGY-----RSLGNVIMCRKGEVALNPLRCKQKR 65
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 157 ELEHGVSTTQRTFKVDIAVYCTAGYTTTKQKGEA--EQANQM-SLTP--QCNKL 211
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 66 PCG-----HPGDDTPGTTTLTGANVEYGVKAVYTCNCGYQLLGSHINYRECDTDG 115
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 212 MCSSLRLIENGXFHPYK-----QYERGEDLVQFCHENNYLSGS-DLIQCNFG 259
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 116 WFNDDIPICE--VYKCLPYTAPENGIKIVSSAMEP-DREYHGAQVRRVYCNNGYKLEDEEM 172
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 260 WYFESTICGRNRKCPPPYPLMSKI-----QHSSTYRHRGEVHIJECETNFVIGSEEL 314
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 173 HCSDDGFMSKREKRCVE---ISCKS-PDVINGSPISOKIYKENERFOYCKMNGEYSE 227
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 315 LC-ENGGKMT-EPKCIIEKEKAVACEQRPSEVENGVAHPHSEIYVSGKVTYRCGGGYSLRG 372
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 228 RDDAVTEGSGWRPLPSCGE--KSC-DNPTIPNG-DYSPLRIKIRKTDDEITYQCRNGFYPA 283
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 373 SSTTCNKRKRWTLPRPCVNIENCKPPPIANGVVVDGLASYTGSVSEYRC-NERYLL 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 284 TRGNIAKCTSTGMIPARCTLKPCDYPDIKH 314
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 432 KGETSRCEGAMSSPPVC-LEPCTI-DVDH 460
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

RESULT 13
LEM3_HUMAN STANDARD; PRT; 830 AA.
AC P16109;
DT 01-APR-1990 (Rel. 14. Created)
DT 01-AUG-1990 (Rel. 15. Last sequence update)
DT 01-OCT-2000 (Rel. 40. Last annotation update)
DE P-SELECTIN PRECURSOR (GRANULE MEMBRANE PROTEIN 140) (GMP-140) (PADGEM)
DE (CD62P) (LEUCOCYTE-ENDOTHELIAL CELL ADHESION MOLECULE 3) (LECAM3).
GN SELP OR GMRP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89168432.
RA Johnston G.I., Cook R.G., McEver R.P.;
RT "Cloning of GMP-140, a granule membrane protein of platelets and
RT endothelium: sequence similarity to proteins involved in cell
RT adhesion and inflammation.";
RT Cell 56:1033-1044(1989).
RN [2]
RP PALMITOYLATION.
RX MEDLINE; 9326599.
RA Fujimoto T., Stroud E., Whalley R.E., Prescott S.M., Muszbek L.,
RA Laposata M., McEver R.P.;
RT "p-selectin is acylated with palmitic acid and stearic acid at
RT cysteine 766 through a thioester linkage.";
RL J. Biol. Chem. 268:11394-11400(1993).
RN [3]
RP STRUCTURE BY NMR OF 160-199.
RX MEDLINE; 97057176.
RA Freedman S.J., Sanford D.G., Bachovchin W.W., Furie B.C., Baleja J.D.,
RA Furie B.;
RT "Structure and function of the epidermal growth factor domain of P-
RT selectin.";
RL Biochemistry 35:13733-13744(1996).
RN [4]
RP 3D-STRUCTURE MODELING OF 42-161.
RX MEDLINE; 94093388.
RA Bajorath J., Stenkamp R., Aruffo A.;
RT "Knowledge-based model building of proteins: concepts and examples.";
RL Protein Sci. 2:1798-1810(1993).
RN [5]
RP VARIANTS ASN-331; ASP-603; VAL-640 AND PRO-756.
RX MEDLINE; 98334547.
RA Hermann S.M., Ricard S., Nicaud V., Mallet C., Evans A.,
RA Ridevets J.B., Arveiler D., Luc G., Cambien F.;
RT "The P-selectin gene is highly polymorphic: reduced frequency of the
RT pro715 allele carriers in patients with myocardial infarction.";
RL Hum. Mol. Genet. 7:1277-1284(1998).
CC -I- FUNCTION: CA(2+)-DEPENDENT RECEPTOR FOR MYELOID CELLS THAT BINDS
CC TO CARBOHYDRATES ON NEUTROPHILS AND MONOCYTES. MEDIATES THE
CC INTERACTION OF ACTIVATED ENDOTHELIAL CELLS OR PLATELETS WITH
CC LEUCOCYTES. THE LIGAND RECOGNIZED IS SIALYL-LEWIS X.
CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -I- TISSUE SPECIFICITY: STORED IN THE ALPHA-GRANULES OF PLATELETS
CC AND WEIBEL-PALADE BODIES OF ENDOTHELIAL CELLS. UPON CELL
CC ACTIVATION BY AGONISTS, P-SELECTIN IS TRANSPORTED RAPIDLY TO
CC THE CELL SURFACE.
CC -I- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC -I- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 9 SUSHI (SCR) REPEATS.
CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD62P entry;
CC WWW="HTTP://WWW.NCBI.NIH.GOV/PROW/CD/CD62P.HTM".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M60234; AAA35910.1; -
DR EMBL; M60217; AAA35910.1; JOINED.
DR EMBL; M60218; AAA35910.1; JOINED.
DR EMBL; M60219; AAA35910.1; JOINED.
DR EMBL; M60222; AAA35910.1; JOINED.
DR EMBL; M60223; AAA35910.1; JOINED.
DR EMBL; M60224; AAA35910.1; JOINED.
DR EMBL; M60225; AAA35910.1; JOINED.
DR EMBL; M60226; AAA35910.1; JOINED.
DR EMBL; M60227; AAA35910.1; JOINED.

```



	FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .)	(POTENTIAL).
	FT	CARBOHYD	716	716	N-LINKED (GLCNAC. . .)	(POTENTIAL).
	FT	CARBOHYD	723	723	N-LINKED (GLCNAC. . .)	(POTENTIAL).
	FT	CARBOHYD	741	741	N-LINKED (GLCNAC. . .)	(POTENTIAL).
	FT	LIPID	807	807	PALMITATE.	
	FT	SITE	818	821	ENDOCYTOSIS SIGNAL (PROBABLE).	
	FT	VARIANT	331	331	S > N.	
	FT	VARIANT	603	603	/FtId=VAR_004192.	
	FT	VARIANT	640	640	N -> D.	
	FT	VARIANT	640	640	/FtId=VAR_004193.	
	FT	VARIANT	756	756	L -> V.	
	FT	VARIANT	756	756	/FtId=VAR_004194.	
	FT	SEQUENCE	830 AA;	90844 MW;	T -> P (REDUCED FREQUENCY IN PATIENTS WITH MYOCARDIAL INFARCTION).	
	FT	SEQUENCE	830 AA;	90844 MW;	/FtId=VAR_004195.	
	FT	SEQUENCE	830 AA;	90844 MW;	FRC407BA2579F6EB CRC64:	
	SO	SEQUENCE	830 AA;	90844 MW;	FRC407BA2579F6EB CRC64:	
	Query Match		15.2%;	Score 286;	DB 1;	Length 830;
	Best Local Similarity		29.6%;	Pred. No. 9.6e-15;		
	Matches	87;	Conservative	38;	Mismatches	135;
					Indels	34;
					Gaps	20;
OY	27	GTQATLYKCRPGYRSLGNVIMWC-RKEGVALNPLRKCKQRPCGHGDP-FSTFTLTGG-	83			
Db	347	GSSCFECQCPGYRVAGLDMLRCIDSGHMSA--PLETCEAISC-EPLESPVHSMCSPSL	403			
OY	84	NVEEGVAAVTYCNGCYQLLEINRECDTDG-WINDPICEVWCLPTAENGKIYS	142			
Db	404	RAFDQDTNCSCFGACGFMLRGADIVR-CENLGQMTAPPAVCALQQDDLPVNEARV--N	460			
OY	143	AMEPDREYHFQGAVERFCNSGYIKIGDEEMHCSDGFSKPKCVETLSCK--SPDVIN	199			
Db	461	CSHPGARIRYGVCSTFCMEGLLVAGASVLQCLATGNNMSYPPECAICTPLTSQ--N	518			
OY	200	GS-----PISQKIYKENERFYKCNMGYEYS--ERGDVACTESG-WRLP-PSCSEKSD	250			
Db	519	GMTCVQPLGSS-SYKST--COFICDEGYSLSGPERLD--CTRSGRWTDSPMCIAIKCP	573			
OY	251	NPYYING---DYSPLRKHRTGDEITTYOORNFYPRTRNTAKNSTG-WIPAP	300			
Db	574	ELFAPEQGSILDSTRKEFPNVGSTCHFSNNGF-KIEGPINVECTYSGRMSATP	626			
RESULT	14					
ID	APOH_MOUSE	STANDARD:	PRT:	345 AA.		
AC	Q01339;					
DT	01-APR-1993 (Rel. 25, Created)					
DT	01-APR-1993 (Rel. 25, Last sequence update)					
DT	01-OCT-2000 (Rel. 40, last annotation update)					
DE	BETA-2-GLYCOPROTEIN I PRECURSOR (APOLIPOPROTEIN H) (APO-H) (B2GP1)					
DN	(BETA(2)GP1) (ACTIVATED PROTEIN C-BINDING PROTEIN) (APC INHIBITOR).					
GN	APOH OR B2GP1.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE: 923372000.					
RA	Nonaka M., Matsuda Y., Shirosishi T., Moriwa K., Natsume-Sakai S.;					
RT	"Molecular cloning of mouse beta 2-glycoprotein I and mapping of the					
RT	gene to chromosome 11."					
RL	Genomics 13:1082-1087(1992).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=CBA/J; TISSUE=LIVER;					
RX	MEDLINE: 94242017.					
RA	Sellar G.C., Steel D.M., Zafiroopoulos A., Seery L.T.,					
RA	Whitehead A.S.;					
RT	"Characterization, expression and evolution of mouse beta 2-					
RT	glycoprotein I (apolipoprotein H)".					
RL	Biochem. Biophys. Res. Commun. 200:1521-1528(1994).					
RN	[3]					



RA	MEDLINE; 95179107 .
RX	Wenzel K., Felix S., Kleber F.X., Brachold R., Menke T., Schatke S.,
RA	Schulte K.L., Glaser C., Rohde K., Baumann G., Speer A.;
RT	"E-selectin polymorphism and atherosclerosis: an association study.";
RL	Hum. Mol. Genet. 3:1935-1937(1994) .
RN	[8]
RP	VARIANT ARG-149.
RX	MEDLINE; 99134508 .
RA	Ye S.O., Usher D., Virgil D., Zhang L.Q., Yochim S.E., Gupta R.;
RT	"A p56l polymorphism detects the mutation of serine-128 to arginine in
RT	J. Clin. Invest. 103:128-134(1999) .
CC	-1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC	MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIAND RECOGNIZED BY
CC	EIAM-1 IS STAIL-Lewis X (ALPHA1->3)FUCOSYLATED DERIVATIVES OF
CC	POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC	GALCTOPLIDS).
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC	-1- POLYMORPHISM: A POLYMORPHISM IN POSITION 149 IS ASSOCIATED WITH A
CC	HIGHER RISK OF CORONARY ARTERY DISEASE (CAD) . A SIGNIFICANTLY
CC	HIGHER MUTATION FREQUENCY (ARG-149) IS OBSERVED IN PATIENTS WITH
CC	ANGIOGRAPHICALLY PROVEN SEVERE ATHEROSCLEROSIS COMPARED WITH AN
CC	UNSELECTED POPULATION (SER-149)
CC	-1- SIMILARITY: TO OTHER SELECTINS/LECAMs.
CC	-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC	-1- SIMILARITY: CONTAINS 6 SUSHI (SCR) REPEATS.
CC	-1- DATABASE: NAME=PROV; NOTE=CD guide CD62e entry;
CC	WWW="HTTP://WWW.NCBI.nlm.nih.gov/prov/cd/cd62e.htm".
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M30640; AAA52377.1; -
DR	EMBL; M61893; AAA52375.1; -
DR	EMBL; M61895; AAA52375.1; JOINED.
DR	EMBL; M61887; AAA52375.1; JOINED.
DR	EMBL; M61888; AAA52375.1; JOINED.
DR	EMBL; M61890; AAA52375.1; JOINED.
DR	EMBL; M61891; AAA52375.1; JOINED.
DR	EMBL; M61892; AAA52375.1; JOINED.
DR	EMBL; M24736; AAA52376.1; -
DR	PIR; A32606; A32606.
DR	PIR; A35046; A35046.
DR	PDB; 1JEA; 03-APR-96.
DR	MIM; 131210; -
DR	INTERPRO: IPR000436; -
DR	INTERPRO: IPR000561; -
DR	INTERPRO: IPR001304; -
DR	INTERPRO: IPR002396; -
DR	PFAM; PF00008; EGF_1.
DR	PFAM; PF00059; Lectin_c.1.
DR	PFAM; PF00084; sushi; 6.
DR	PRINTS; PR00343; SELECTIN.
DR	PROSITE; PS00022; EGF_1; 1.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS00615; C-type_Lectin_2; 1.
KM	Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KM	Selectin; Signal; Sushi; Repeat; Polymorphism; 3D-structure.
FT	SIGNAL 1 21
FT	CHAIN 22 610
FT	DOMAIN 22 556 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 557 578 POTENTIAL.
FT	DOMAIN 579 610 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 38 138 C-TYPE LECTIN (SHORT FORM).

FT	DOMAIN	139	175	EGF-LIKE.	
FT	DOMAIN	179	548	6 X SUSHI (SCR)	REPEATS.
FT	REPEAT	179	238	SUSHI 1.	
FT	REPEAT	241	300	SUSHI 2.	
FT	REPEAT	303	363	SUSHI 3.	
FT	REPEAT	366	426	SUSHI 4.	
FT	REPEAT	429	489	SUSHI 5.	
FT	REPEAT	492	548	SUSHI 6.	
FT	DISULFID	40	138		
FT	DISULFID	111	130		
FT	DISULFID	143	154		
FT	DISULFID	148	163		
FT	DISULFID	165	174		
FT	DISULFID	180	224		
FT	DISULFID	210	237	BY SIMILARITY.	
FT	DISULFID	242	286	BY SIMILARITY.	
FT	DISULFID	272	299	BY SIMILARITY.	
FT	DISULFID	304	349	BY SIMILARITY.	
FT	DISULFID	335	362	BY SIMILARITY.	
FT	DISULFID	367	412	BY SIMILARITY.	
FT	DISULFID	398	425	BY SIMILARITY.	
FT	DISULFID	430	475	BY SIMILARITY.	
FT	DISULFID	461	488	BY SIMILARITY.	
FT	DISULFID	493	534	BY SIMILARITY.	
FT	DISULFID	520	547	BY SIMILARITY.	
FT	CARBOHYD	25	25	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	160	160	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	179	179	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	199	199	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	203	203	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	265	265	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	312	312	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	332	332	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	503	503	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	527	527	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARIANT	149	149	S -> R (ASSOCIATED WITH A RISK FACTOR FOR CAD).	
FT				/FTID-VAR 004191.	
FT				H -> Y (TN REF. 2)	
SEQ	CONFLICT SEQUENCE	468 610 AA;	468 66655 MM;	7D43E3C0D1229229	CRC64:

[illegible]

Tue Nov 21 16:57:30 2000

us-09-316-163-11.rsp

Page 18

Search completed: November 21, 2000, 16:49:24  
Job time: 327 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:51 ; Search time 70.54 Seconds  
(without alignments)  
435.499 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDCNLPERRRNTTEITGSMS.....PDIKHGLYHEMRPRFPV 329

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTREMBL\_14:\*  
2: sp.archaea:\*  
3: sp.bacteria:\*  
4: sp.fungi:\*  
5: sp.human:\*  
6: sp.invertebrate:\*  
7: sp.mammal:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1876	100.0	449	4	014570
2	903	48.1	669	6	028085
3	486	25.9	1053	13	091275
4	377	20.1	522	6	028769
5	353	18.8	679	11	099254
6	351	18.7	1911	6	029528
7	350.5	18.7	559	4	090V2
8	350.5	18.7	2039	4	016745
9	350.5	18.7	2489	4	016744
10	349.5	18.6	645	12	09WPU2
11	348	18.6	2014	6	029530
12	342.5	18.3	661	6	029531
13	339	18.1	522	6	028797
14	337.5	18.0	560	5	022328
15	335	17.9	360	12	09YU08
16	333.5	17.8	315	6	028770
17	328	17.5	550	12	P88903
18	328	17.5	550	12	040912
19	323	17.2	497	11	063612

20	323	17.2	559	11	063135	063135 rattus norv
21	317	16.9	452	11	061407	061407 mus musculu
22	310	16.5	259	12	P87616	P87616 compox viru
23	308.5	16.4	1032	4	013866	013866 homo sapien
24	306.5	16.3	417	11	035520	035520 rattus norv
25	306	16.3	303	11	061405	061405 mus musculu
26	304	16.2	483	11	064735	064735 mus musculu
27	302.5	16.1	657	4	014006	014006 homo sapien
28	302	16.1	533	11	008569	008569 cavia porce
29	298	15.9	1124	5	09YU4	09YU4 drosophila
30	294	15.7	1652	5	09VU09	09VU09 drosophila
31	293	15.6	363	6	002839	002839 sus scrofa
32	292.5	15.6	1045	6	046545	046545 ovis aries
33	290	15.5	1087	4	014212	014212 homo sapien
34	285	15.2	579	11	060736	060736 mus musculu
35	284.5	15.2	263	12	089859	089859 variola vir
36	284.5	15.2	263	12	007033	007033 variola vir
37	284	15.1	740	4	095508	095508 homo sapien
38	283.5	15.1	263	12	089076	089076 variola vir
39	283.5	15.1	974	5	P91658	P91658 drosophila
40	280.5	15.0	263	12	089061	089061 variola vir
41	278	14.8	354	5	090611	090611 drosophila
42	274.5	14.6	349	4	015429	015429 homo sapien
43	272	14.5	285	6	019126	019126 macaca fasc
44	272	14.5	285	6	019127	019127 macaca mula
45	270.5	14.4	369	6	P79138	P79138 cercopithec

## ALIGNMENTS

RESULT 1

ID 014570 PRELIMINARY: PRT: 449 AA.

AC 014570: P78435;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE COMPLEMENT FACTOR H PRECURSOR.

GN HF OR CFH.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 88134059.

RA Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;

RT "The complete amino acid sequence of human complement factor H.";

RL Biochem. J. 249:593-602(1988).

RN [2]

RP SEQUENCE OF 226-449 FROM N.A.

RX MEDLINE: 86169701.

RA Kristensen T., Wetsel R.A., Tack B.F.;

RT "Structural analysis of human complement protein H: homology with C4b binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";

RL J. Immunol. 136:3407-3411(1986).

RN [3]

RP SEQUENCE OF 1-19 FROM N.A.

RA Vik D.P., Williams S.A.;

RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 1-9 FROM N.A.

RA Dominguez O.;

RT Thesis (1993), Immunologia, Hospital Trias I Pujol, Spain.

RL EMBL: X07523; CAA30403.1; -.

DR EMBL: M12383; AAA52013.1; -.

DR EMBL: U56979; AAB01987.1; -.

DR EMBL: Z29655; CAA82763.1; -.

DR HSSP: P10998; IYVC.

DR INTERPRO: IPR000436; -.

DR PIRAM: PF000084; sushi; 7.

KW SIGNAL.

FT SIGNAL 1 18 POTENTIAL.

FT CHAIN 19 449 POTENTIAL.  
SQ SEQUENCE 449 AA: 51033 MW: C2AAD4/F155343E3 CRC64;

Query Match 100.0%; Score 1876; DB 4; Length 449;  
Best Local Similarity 100.0%; Pred. No. 6.8e-161;  
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPPRRNTLILGMSWSDQTYPECTQAIYKCRPGYBSLGNVIMVCRGGEVVALPLR 60  
DB 19 EDCNELPPRRNTLILGMSWSDQTYPECTQAIYKCRPGYBSLGNVIMVCRGGEVVALPLR 78  
QY 61 KCGKRCGHHGDPFGFTLTLGNVFEYGVKAYVTCNEGQQLGELINRCDTDGWTNDI 120  
DB 79 KCGKRCGHHGDPFGFTLTLGNVFEYGVKAYVTCNEGQQLGELINRCDTDGWTNDI 138  
QY 121 PICEVVKCLPVTAPENGKIYSSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFV 180  
DB 139 PICEVVKCLPVTAPENGKIYSSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFV 198  
QY 181 SKKPKCVELSCSPDVINGSPISOKIYKENRFPQYKCMGYEYSEKDAVCTESGWRP 240  
DB 199 SKKPKCVELSCSPDVINGSPISOKIYKENRFPQYKCMGYEYSEKDAVCTESGWRP 258  
QY 241 LPSCEKSCDNPYIPNGDYSPLRIKHTGDEITYOCNNGFYPATRGNTACTSTGWTIPAP 300  
DB 259 LPSCEKSCDNPYIPNGDYSPLRIKHTGDEITYOCNNGFYPATRGNTACTSTGWTIPAP 318  
QY 301 KCTLPKCDYPDIKHGLYHENMRPFPV 329  
DB 319 KCTLPKCDYPDIKHGLYHENMRPFPV 347

RESULT 2  
Q28085 PRELIMINARY: PRT; 669 AA.  
ID Q28085  
AC Q28085;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
DE CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 96202005.  
RA Soames C.J., Day A.J., Sim R.B.;  
RT "Prediction from sequence comparisons of residues of factor H involved  
in the interaction from sequence complement component C3b.";  
RL Biochem. J. 315:523-531(1996).  
DR EMBL: X98697; CA67257.1; -;  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -;  
DR PFAM: PF00084; sush1.11.  
FT NON\_TER 1  
FT NON\_TER 669  
SQ SEQUENCE 669 AA: 75683 MW: DDD9DB30E747AC2 CRC64;

Query Match 48.1%; Score 903; DB 6; Length 669;  
Best Local Similarity 60.9%; Pred. No. 3.8e-73;  
Matches 154; Conservative 33; Mismatches 64; Indels 2; Gaps 1;

QY 76 GTFTLTGNGVEYGVKAYVTCNEGQQLGELINRCDTDGWTNDIPICEVVKCLPVTAP 135  
DB 3 GSPHAEENOFEGYGAKYVTCDEGYQWGVNMFRECDTNGMTNDIPICEVVKCLPVTAP 62  
QY 136 NGKITSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFVSKKPKCVELSCSP 195  
DB 63 NGKITSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFVSKKPKCVELSCSP 122

QY 196 DVINGSPISOKIYKENRFPQYKCMGYEYSEKDAVCTESGWRPPLPSCEKSCDNPYIP 255  
DB 123 VILNGAVLPKATYKONEROYRCAAGFEGGRDVTCTSGMPAPATCTEITDDPRIP 182  
QY 256 NGDYSPLRIKHTGDEITYOCNNGFYPATRGNTACTSTGWTIPAPCTLPKCDYPDIKH 315  
DB 183 NGVIRPELSKYRGODKITYEKKKGFFPEIRGTATCTRDGVVPVRCAMKPCSYPIKHG 242  
QY 316 GLYHENMRPFPV 328  
DB 243 RLYYS--YRGYF 253

RESULT 3  
Q91275 PRELIMINARY: PRT; 1053 AA.  
ID Q91275  
AC Q91275;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
DE COMPLEMENT REGULATORY PLASMA PROTEIN.  
OS Paratubrax nebulifer (barred sand bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;  
OC Perciformes; Percoidae; Serranidae; Paratubrax.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RX MEDLINE: 94318039.  
RA Dahmen A., Kaidoh T., Zipfel P.F., Gigli I.;  
RT "Cloning and characterization of a cDNA representing a putative  
complement-regulatory plasma protein from barred sand bass (Paratubrax  
nebulifer)."  
RL Biochem. J. 301:391-397(1994).  
DR EMBL: L21703; AAA92556.1; -;  
DR HSSP: P08603; IHPH.  
DR INTERPRO: IPR000436; -;  
DR PFAM: PF00084; sush1.16.  
SQ SEQUENCE 1053 AA: 117597 MW: F27E32C3AD76D5D3 CRC64;

Query Match 25.9%; Score 486; DB 13; Length 1053;  
Best Local Similarity 29.7%; Pred. No. 2.1e-35;  
Matches 114; Conservative 48; Mismatches 134; Indels 88; Gaps 13;

QY 21 DQTYPEGTQAIYKCRPGYBSLGNVIMVCRGGEVVALPLRCKCRPGHGDTPFGFTL 80  
DB 47 EASYPGRGVRCVNCVGS--GFFKLVCEBGKMETG--AKQPRSCGHHGDAQFADPHL 102  
QY 81 TGNVFEYGVKAYVTCNEGQQLGELINRCDTDGWTNDIPICEVVKCLPVTAPENGKIV 140  
DB 103 AECNDFVFGSKVYVTCQKGYOMVSRINRYRCVAEGDGVVPCESQOCPLIHNDNVQYI 162  
QY 141 SSAMEPDRREYHFGQAVRFVNSGKIEGDEMHCSDDGFVSKKPKCVELSCSPDVIN 199  
DB 163 GG-----PEATGNGVNVRFSCKSRELSDSPELYCDERGDWSPVYKCAITCALPIEN 218  
QY 200 GSPISOKIYKENRFPQYKCMGYEYSEKDAVCTESG-----WRPLPSCEKSC----- 249  
DB 219 GNVPGAIRREKENDVLYHCDRAFKHIDR--PSTCIQGIKAEMSPPLDESJKRRTIWD 277  
QY 250 -----DNPIIP-----NGDYS----- 260  
DB 278 GTRYEPAYRNLPSPGFTLVICARTSWISTPQETSIVTTCODNGEWSIRPTQOEVRCSNR 337  
QY 261 -PLRI-----KHTGDEITYOCNNGFYPATRGNT--AKTSTGWTIPAPCTLPKCDY 309  
DB 338 REHDSMWVRSMERYTLDNDRYMKRG--YKRTGSGVTATCGRNGMMPNPLCEVYKTSK 396  
QY 310 PDIKHG-----GLYHENMRPFPV 326  
DB 397 ENIDQAVIVGTDKQIYNLNOKALY 420

RESULT 4  
ID 028769 PRELIMINARY; PRT: 522 AA.  
AC 028769;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE COMPLEMENT RECEPTOR (FRAGMENT).  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Papio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BONE MARROW;  
RA Blumhagen D.J., Logar C.M., Shen X.P., Chen W.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: L77977; AAA93004.1; -.  
DR HSSP: P08603; IHT1.  
DR INTERPRO: IPR000436; -.  
DR PFAM: PF00084; sushi; 7.  
FT NON\_TER 1  
SQ SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

Query Match 20.1%; Score 377; DB 6; Length 522;  
Best Local Similarity 30.2%; Pred. No. 5.8e-26;  
Matches 107; Conservative 55; Mismatches 142; Indels 50; Gaps 22;

QY 3 CN---ELPPRTELITLSSWSDQTYPECTQAIYKCRPGYRSLGNVIMCRKGEWYALNPL 59  
DB 48 CNAEQLPFPARTNLTDS---EPFVGTLYKECLPGHGKRFSLICLKNVMTSADK- 102  
QY 60 RKCKRPGCHGDPFGFTLTGNGVAVYKAVYTCNGYOLLGEINREDTDC--- 115  
DB 103 -KCTKSKRNKRDYPNGAVHY--KDIOFGSIOYNSCKNGYLLIS--SATCTIISGNYI 158  
QY 116 WTNDIPICEVYKC-LPVTAPENGKIVSSAMEPDRYH-FGQAVRFVYVNSG-----YKLE 167  
DB 159 WDNETPICEIIPGGLPPTI-ANGDFISTSS---REYFPGSVVYTRCNLSGRKKLELV 213  
QY 168 GDEEMHCS--DD--GFWSKERPKC-VEISCKSPDVINGSPIS-QKIYKENERPOYCKM 221  
DB 214 GEPSTYCSKSDQVIGWGPAPQCIIPKCMPPNENGLVSNVSLFSLNVEFERCP 273  
QY 222 GYEYSEKSDAVCTE-SGWRP-LPSCSEKSCDNPYIPNGDYSPLRIKRTGDEITYQCRRG 279  
DB 274 GFVMGPRHVOQALNKKEPELPSRCVQPPPELHGHHTSHODFSGQEVYFSCPEG 333  
QY 280 FYPATRGNTAKCTSTG-WIP-APRCTLKPCD--YPDIKHGGLYHNMRRPYFPV 329  
DB 334 -YDLRGAASLHCTPGQDWNPEAPICTYKSCDDFLGLPHG-----FVLPPL 378

RESULT 5  
ID 099254 PRELIMINARY; PRT: 679 AA.  
AC 099254;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE COMPLEMENT RECEPTOR TYPE 2 (CR2) (FRAGMENT).  
GN CR2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE: 90229754.  
RA Kurtz C.B., O'Toole E., Christensen S.M., Weis J.H.;

RT "The murine complement receptor gene family. IV. Alternative splicing  
RT of Cr2 gene transcripts predicts two distinct gene products that share  
RT homologous domains with both human CR2 and CR1."  
RL J. Immunol. 144:3581-3591(1990).  
RN [2]  
RP SEQUENCE OF 21-367 FROM N.A.  
RX MEDLINE: 95105691.  
RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seta T., Wagner L.M.,  
Hollers V.M.;  
RT "Mouse complement regulatory protein Cr2/p65 uses the specific  
RT mechanisms of both human decay-accelerating factor and membrane  
RT cofactor protein."  
RL J. Exp. Med. 181:151-159(1995).  
DR EMBL: U17128; AAA78271.1; -.  
DR EMBL: U17123; AAA78271.1; JOINED.  
DR EMBL: U17124; AAA78271.1; JOINED.  
DR EMBL: U17125; AAA78271.1; JOINED.  
DR EMBL: U17126; AAA78271.1; JOINED.  
DR EMBL: U17127; AAA78271.1; JOINED.  
DR EMBL: M36470; AAA37449.1; -.  
DR HSSP: P10998; IYVC.  
DR MGD: MGI:88489; Cr2.  
DR INTERPRO: IPR000436; -.  
DR PFAM: PF00084; sushi; 10.  
FT NON\_TER 679  
SQ SEQUENCE 679 AA; 74916 MW; 52FC00FDCED20CDC CRC64;

Query Match 18.8%; Score 353; DB 11; Length 679;  
Best Local Similarity 28.9%; Pred. No. 1.1e-23;  
Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

QY 20 SPQT-YPECTQAIYKCRPGYRSLGNVIMCRKGEWYALNPLRKCKRPGCHGDPFGFT 78  
DB 37 SDKSEFALGTETWEYKCRPFRYKRSFTICTLETSSK--SDAQOFCKRRKPCMNQEPHGSV 94  
QY 79 TLGNGVAVFYKAVYTCNGYOLLGEINREDTDC---DPDGMWNTDIPICEVYKCLPVTAP 134  
DB 95 HINTG--IEFGSTITYSCNGGTRLIQD--SATCTIYSDNTVMNDMLCESIPCESPPAI 151  
QY 135 ENKIVSSAMEPDRYHFGQAVRFVYVNSG-----YKIGDEEMHCSDD-----GFWSKER 184  
DB 152 SNGDFYSSSRD---SFFYGMVYTYCHTGKRNREKLFDLVGEKICYTSDNGVINSNP 208  
QY 185 PKCV-EISCKSPDVINGSPISQ-KIYKENERPOYCKMNGYESEKDAVC-TESGWR-P 240  
DB 209 POCIPRVKCPMEIENGELVSEGFKHSFPLNDYVIFKCKSGFTYKKSRIAMQPNKSWSP 268  
QY 241 LPSCSEKSCDNPYIPNGDYSPLRIKRTGDEITYQCRRGFPYATRGNTAKCTSTG-WI-P 298  
DB 269 LPTCFMGCLPPONLIHGDYNNKDKDEFVSQOKVSYTCNPG-YTLGLNVECTSLGTSMT 327  
QY 299 APRCTLKPCD-YPD-IRHGGLY 318  
DB 328 VPTCEYKSCDAIPNHLHGRVF 349

RESULT 6  
ID 029528 PRELIMINARY; PRT: 1911 AA.  
AC 029528;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE COMPLEMENT RECEPTOR 1 (FRAGMENT).  
GN CR1.  
OS Papio hamadryas (Hamadryas baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Papio.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.;

RA Atkinson J.P.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L39791; AAA62170.1; -.  
 DR HSSP; P08603; IHCC.  
 DR INTERPRO; IPR000436; -.  
 DR INTERPRO; IPR001424; -.  
 DR INTERPRO; IPR000834; -.  
 DR PPM; PF00084; sushi; 29.  
 DR PROSITE; PS00087; SOD\_CU\_ZN\_1; UNKNOWN\_1.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_4.  
 FT NON\_TER 1911 1911  
 SQ SEQUENCE 1911 AA: 210173 MW: 535AADD0EAF521D CRC64:

Query Match 18.7%; Score 351; DB 6; Length 1911;  
 Best Local Similarity 29.0%; Pred. No. 6,1e-23;  
 Matches 93; Conservative 57; Mismatches 135; Indels 36; Gaps 21;

QY 20 SDQTYEGTQAIYKCRGYSLSGNVIMV-CRKG-EMVALPRLKCKQKRCGHPGDTTFFGT 77  
 Db 1299 NDEFPYGSILNECHPGY--FGMRFSISCLEMLWMSVSD--NCRKSCGTPE--PFNG 1353  
 QY 78 FTLLGNGVEYGYKAVYTCNEGYQLLEINVRCDTGD---WTNDIPICEVVKCLEPVA 133  
 Db 1354 MVAHINTDT-QFGSTVNSCNEGFRLLISPS-TTCIVSGNNVMTDKAPICEIITISCKRPPT 1411  
 QY 134 PENGKIVSSAMEDREYHFGQAVFVCSNG-----YKIGDEEMHCS--DD--GFWSKE 183  
 Db 1412 ISNGDEYSNNR---TSRSGTAVTYOCHTGPDDEQLFELVGERSIYCTSKDDYGAWSSP 1468  
 QY 184 KPKCEVLS-CKSPDYINGSPI-SOKIITYKENERFOYKCNNGYSESGDAVC--TESGMRP 240  
 Db 1469 PPRCISTNCKTAPEVNGNIVPGRNRSFSLNELYFRFCQPEFVWGSHTYOCOTNNRNGP 1528  
 QY 241 -LPSCEKSCDNPIYNGDISPL-RIKHRTGDEITYQCNGFYPAIRGNTAKCTSTG-WT 297  
 Db 1529 KLPHCSRVCPPELILHGETPISHDKFSPGOEVFVSCDEPG-YDLRGAASLHCTPGQDWS 1587  
 QY 298 P-APRCTLKPCD--YPDIKHG 315  
 Db 1588 PEAPICTVKSDEFLGQLPRHG 1608

RESULT 7  
 Q9UOV2 PRELIMINARY: PRT; 559 AA.  
 AC Q9UOV2;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
 DE HUMAN CRI MRNA FOR C3B/C4B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89010527.  
 RA Houchard D., Miesner D.R., Atkinson J.P., Holers V.M.;  
 RT "Identification of an alternative polyadenylation site in the human  
 RT C3b/C4b receptor (complement receptor type 1) transcriptional unit and  
 RT prediction of a secreted form of complement receptor type 1.";  
 RL J. Exp. Med. 168:1255-1270(1988).  
 DR EMBL; X14362; CAA32541.1; -.  
 DR INTERPRO; IPR000436; -.  
 DR INTERPRO; IPR000834; -.  
 DR PPM; PF00084; sushi; 8.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 KW Signal.  
 FT NON\_TER <1 1  
 FT SIGNAL 17 559 POTENTIAL.  
 FT CHAIN 17 559  
 SQ SEQUENCE 559 AA: 61424 MW: DBFPE965CA179D75 CRC64:

Query Match 18.7%; Score 350.5; DB 4; Length 559;  
 Best Local Similarity 29.3%; Pred. No. 1,5e-23;  
 Matches 100; Conservative 55; Mismatches 127; Indels 59; Gaps 23;

QY 3 CNE---LPRRNTEILTGSWSDQTYPRGTOAIYKCRGYSLSGNVIMVCRKGEWVALNPL 59  
 Db 18 CNAPEWLPFAFAPMTLT---DEFEFFPIGTLYNRCRGYSGRPSITLCKNSVYTGAKD- 72  
 QY 60 RKCKRPGCGHPGDPPTFTLLTGNVEYGYKAVYTCNEGYQLLEINVRCDTGD---PTDG 115  
 Db 73 -RCRRKSCRNPPDPVNGMVAHIK--TQFSQILYSCYKGRILGS--SARCTIISGPTVI 128  
 QY 116 WTNDIPICEVVK-LPYTAENGKIVSSAMEDREYHFGQAVFVCSNG-----YKTEG 168  
 Db 129 WDNETPTICDRIPGLPPTI-TNGDFTSTNR---NFHGSVITYRCNPGSGGRVFEIYG 184  
 QY 169 DEEMHC--SD--GFWSKEPKC-VEISCKSPDYINGSPI-SOKIITYKENERFOYKCNMG 222  
 Db 185 EPSIYCTSNDDQYIGSGPAPQCIIPKRCPPPVNENGLVSDNLSFLNEVVEFRCPG 244  
 QY 223 YEYSEKGDVCTE-SGRP-LPSCEKSCDNPIYPN-----GDYSPLRIKHRTG 269  
 Db 245 FVAKGPRRVQALNKEPELPSC-SRVQCP--PDVLAEKTRQDKDNFSP-----G 294  
 QY 270 DEITYQCRNGFYPATRGNTAKCTSTG-WIP-APRCTLKPCD 308  
 Db 295 QEVFYSCDEPG-YDLRGAASMRCTPGQDWSAPAPTCVEKSCD 334

RESULT 8  
 Q16745 PRELIMINARY: PRT; 2039 AA.  
 AC Q16745;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE COMPLEMENT RECEPTOR 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94065175.  
 RA Vik D.P., Wong W.W.;  
 RT "Structure of the gene for the F allele of complement receptor type 1  
 RT and sequence of the coding region unique to the S allele.";  
 RL J. Immunol. 151:6214-6224(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Vik D.P., Wong W.W.;  
 RL Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; L17418; AAB60694.1; -.  
 DR EMBL; L17390; AAB60694.1; JOINED.  
 DR EMBL; L17391; AAB60694.1; JOINED.  
 DR EMBL; L17392; AAB60694.1; JOINED.  
 DR EMBL; L17393; AAB60694.1; JOINED.  
 DR EMBL; L17394; AAB60694.1; JOINED.  
 DR EMBL; L17395; AAB60694.1; JOINED.  
 DR EMBL; L17396; AAB60694.1; JOINED.  
 DR EMBL; L17397; AAB60694.1; JOINED.  
 DR EMBL; L17398; AAB60694.1; JOINED.  
 DR EMBL; L17399; AAB60694.1; JOINED.  
 DR EMBL; L17400; AAB60694.1; JOINED.  
 DR EMBL; L17401; AAB60694.1; JOINED.  
 DR EMBL; L17402; AAB60694.1; JOINED.  
 DR EMBL; L17403; AAB60694.1; JOINED.  
 DR EMBL; L17404; AAB60694.1; JOINED.  
 DR EMBL; L17405; AAB60694.1; JOINED.  
 DR EMBL; L17406; AAB60694.1; JOINED.  
 DR EMBL; L17407; AAB60694.1; JOINED.  
 DR EMBL; L17408; AAB60694.1; JOINED.  
 DR EMBL; L17409; AAB60694.1; JOINED.







Query Match 18.3%; Score 342.5; DB 6; Length 661;  
 Best Local Similarity 29.0%; Pred. No. 9.8e-23;  
 Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;

RL J. Immunol. 153:691-700(1994).  
 DR EMBL; L24921: AAA51439.1; -;  
 DR HSSP; P10998; 1VVC;  
 DR INTERPRO; IPR000436; -;  
 DR INTERPRO; IPR001424; -;  
 DR PFAM; PF00084: sush1; 9.  
 DR PROSITE; PS00087: SOD\_CU\_ZN\_1; UNKNOWN\_1.  
 KW Signal; Alternative splicing.  
 FT NON\_TER 1  
 FT SIGNAL <1 16 POTENTIAL.  
 FT CHAIN 17 >661 COMPLEMENT RECEPTOR 1.  
 FT NON\_TER 661 661  
 SQ SEQUENCE 661 AA; 72966 MW; 3F0DF1532664D3F0 CRC64;

Query Match 18.3%; Score 342.5; DB 6; Length 661;  
 Best Local Similarity 29.0%; Pred. No. 9.8e-23;  
 Matches 99; Conservative 55; Mismatches 128; Indels 59; Gaps 23;

QY 3 CNE---LPPRNTEILTGSMSDQTPYECTQAIYKCRPGYSLGNVIMVCKRGEVVALNPL 59  
 DB 18 CNAPEWLPFARPTNL---DEFEFPITGYLNTCECRPGYGRPSIICLKNSVMTGAKD- 72  
 QY 60 KCOGRPCGHPDPTFTLTGTVNFEYGVKAYTCNEGVLGELINRECD---PTDG 115  
 DB 73 -RCRRKCRNPDPVNGMVHT--KDIOFGSQIKYCTKGRILIGS--SSATCIISGDTVI 128  
 QY 116 WTNDIPICEVVKC-LPTVAPENGKIVSSAMEPREHFGAVRVCNSG-----YKIEG 168  
 DB 129 WDNETPICDRIPGLPPTI--TNGDFISTNRE---NFHYGSVVTYRCNCGSGRKVFELVG 184  
 QY 169 DEEMHC--SDD--GFMSKERKPC-VEISCKSPDVINSPTIS-QKIIKENRFPQKCMG 222  
 DB 185 EPSTICTSNDQVGVSGPAPQCIIPKCTPPNVENGILVSDNSLPSLNVFEFCQPG 244  
 QY 223 YVESERDAVCTE-SGNRP-LPSCSEKSCDNPYIPN-----GDYSPLRKIHRTG 269  
 DB 245 FVMKGPVRVCKQALNKKEPELPSG-SRVCCPP--PDVIAHARTQRDKNFSP-----G 294  
 QY 270 DEITYGCRNGFYPATRGNTAKCTSTG-WIPA-PRCTLKPCD 308  
 DB 295 QEVFYSCPEG-YDLRGAASLRCTPGQDWSPATPCVEVSCD 334

RESULT 13  
 ID Q28797 PRELIMINARY; PRT; 522 AA.  
 AC Q28797;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE UNKNOWN PROTEIN (FRAGMENT).  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94292799.  
 RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;  
 RT "Primary sequence of an alternatively spliced form of CrI. Candidate  
 RT for the 75,000 M(r) complement receptor expressed on chimpanzee  
 RT erythrocytes.";  
 RL J. Immunol. 153:691-700(1994).  
 DR EMBL; L24922: AAA50460.1; -;  
 DR HSSP; P08603; 1HFI.  
 DR INTERPRO; IPR000436; -;  
 DR INTERPRO; IPR000834; -;  
 DR PFAM; PF00084: sush1; 8.  
 DR PROSITE; PS00133: CARBOXYPEPT\_ZN\_2; UNKNOWN\_1.  
 FT NON\_TER 1  
 SQ SEQUENCE 522 AA; 57304 MW; DB8456EFA62C00C9 CRC64;

Query Match 18.1%; Score 339; DB 6; Length 522;  
 Best Local Similarity 30.1%; Pred. No. 1.5e-22;  
 Matches 95; Conservative 52; Mismatches 119; Indels 50; Gaps 21;

QY 24 YPECTQAIYKCRPGYRSLGNVIMVCKRGEVVALNPLKCRKRRPGHPDPTFTLTG 83  
 DB 5 FPITTYLNTCECRPGYGRPSIICLKNSVMTSAD--KCRKSCRNPDPVNGMVHT-- 60  
 QY 84 NFVEYGVKAVTYTCNEGVLGELINRECDTDG---WTNDIPICEVVKC-LPTVAPENGK 138  
 DB 61 KDIOFGSQIKYCTKGRILIGS--SSATCIISGNVIMVIMNKPVCDDRIICGLPPTI-ANGD 118  
 QY 139 IVSSAMEPRDREY-HFGAVRVCNSG-----YKIEGDEMHCS--DD--GFMSKERKPC 187  
 DB 119 FTSIS---REYFYASVVTYHCHNCGSGKRVFELVGEPTICTSKDQVGVSGPAPQC 174  
 QY 188 -VEISCKSPDVINSPTIS-QKIIKENRFPQKCMGEYSEERDAVCTE-SGNRP-LPS 243  
 DB 175 IIPKCTPPNVENGILVSDNSLPSLNVFEFCQPGFVMKGPVRVCKQALNKKEPELPS 234  
 QY 244 CEKSCDNPYIPND-----YSPLRKIHRTGDEITYGCRNGFYPATRGNT-AKCTST 294  
 DB 235 CSRVCCPPPDVIAHARTQRDKNFSP-----GDEVYSCPEG--DLRSTYLIHCTPQ 285  
 QY 295 G-WIP-APRCTLKPCD 308  
 DB 286 GDWSPEAPRCEVSCD 301

RESULT 14  
 ID Q22328 PRELIMINARY; PRT; 560 AA.  
 AC Q22328;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE COSMID T07H6.  
 GN T07H6.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 NC Rhabditidae; Peleoderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latelle P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX Geisel C.;  
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX Waterston R.;  
 RA Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53344: AAA96225.1; -;  
 DR HSSP; P10998; 1VVC.  
 DR INTERPRO; IPR000436; -;  
 DR PFAM; PF00084: sush1; 7.  
 SQ SEQUENCE 560 AA; 61619 MW; FB8923BAC1B320C9 CRC64;

Query Match 18.0%; Score 337.5; DB 5; Length 560;  
 Best Local Similarity 23.0%; Pred. No. 2.3e-22;  
 Matches 100; Conservative 57; Mismatches 142; Indels 135; Gaps 15;

```

QY 3 CNELPFRRTTEITGSMWQTYEGTQAIYKCPGRSGLNVMVCRKG-EMVALNPL-- 59
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 17 CNELPFRRTTEITGSMWQTYEGTQAIYKCPGRSGLNVMVCRKG-EMVALNPL-- 59
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 -----RCKCRKPCG 68
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 73 LCLKGPEPEIPFAVHDSFSFSEGYDLDAEGLSISKCLNKKVNAQWFGDLCKRACR 132
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 HGGDIPFGFTLTGCVNFEYGYKAVYTCNBYOLGEINRECDTG-WTNDIPICEYK 127
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 133 DPGDLENG---LREGDTFEYPHHVKYSNCPNFGFLVGTST-RQSSNGEWTEPANCKATE 188
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 128 CLPVTAPENGKIYSSAMEEDREYHFGQAVFVONSGYKIEGDEMHGSDGFEWKEKPC 187
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 189 CSRPSPLHGKRYGSSLT-----YQSVVTTYSCDHGRILGVGVORICLAEIWMGNEPPRC 242
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 188 VEICK-SPDVING-----RCKCRKPCG 203
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 243 EIRCSVLTPLNGLYEGSETSFGAVAVPRCLETMTHEGASKAKCMEDGQMSAPIPRCLA 302
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 204 SOKIYKENERFOYK-----CNMGYFSERGDAVCTESGMRPLPSCEEKS 248
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 303 SCRVPRIQNGKIRDSSEGLIAGSKYIVECKNQHANIDERTICNSTWSHVPCSPLS 362
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 249 CDN--PYIPNGDYSPLRIKHTGDEITYOCRNFGFYPATRGNTAKTSTGWI---PAPRCT 303
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 CHNWPPRVPHARI--LFSKSSHGSIKAYECNNGNHPNRNNQIILKILXGEMTKDGPMPKCL 420
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 304 LKPCDYPDIKHGL 317
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 421 PSCCHPSKTYGTL 434
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 15

```

Q9YTO8 PRELIMINARY; PRT; 360 AA.
AC 09YTO8;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE COMPLEMENT CONTROL PROTEIN HOMOLOG CCPH.
OS Ateline herpesvirus 3.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-73;
RA Albrecht J.-C., Fleckenstein B.;
RT "Primary Structure of the Herpesvirus Atelles Genome.";
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083424; AAC95530.1; -.
DR HSSP; P10998; IYVC.
DR INTERPRO; IPR000436; -.
DR PRAM; PF00084; sush1; 4.
SQ SEQUENCE 360 AA; 40208 MW; 118CF83C034352A0 CRC64;

```

Query Match 17.9%; Score 335; DB 12; Length 360;  
 Best Local Similarity 33.4%; Pred. No. 2.2e-22;  
 Matches 103; Conservative 32; Mismatches 119; Indels 54; Gaps 20;

```

QY 8 PRRTET-----LTGSMWQTYEGTQAIYKCPGRSGLNVMVCRKGEMVALNPL 59
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 24 PKRNRVSLRYVNITN--SSGSPNGTTLQVYCRKGY--IGROIQTVCVNGNMTVPN-- 77
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 RCKCRKPCGHGDPFGFTLTGCVNFEYGYKAVYTCNBYOLGEINRECDTG--DTDG-- 115
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

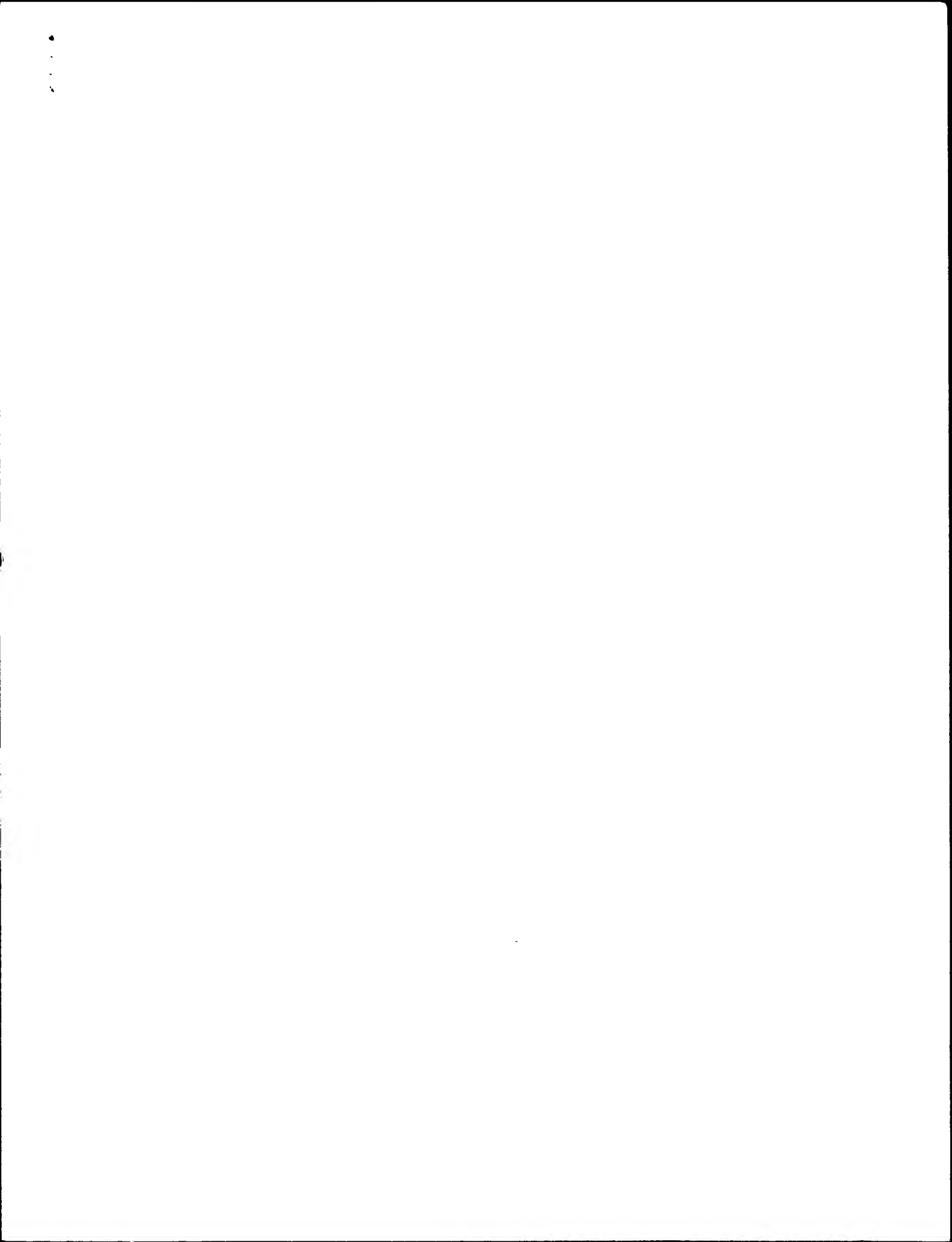
```

DB 78 -EQCKRRCSTPADLLNGWYTVT-GNLY-YGSVITYTCNTGYOLLGSPT-SSCLLGPGRV 133
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 116 -WTNDIPICEYKCLPVTAPENGKIYSSAMEEDREYH-FGQAVFVONSGYK--IGDEE 171
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 134 NMTFRPPICITTCCKRPPPTIANSTHINI-----KEYTYIDAATYISCNDETCLTLTGPSS 188
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 172 MHCSDGFW-SKEKPRCWEISCKSPDVINGSPISOKIYKENERFOY---KCMAGYEYSE 227
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 189 KQCSSETRRWVPDEETKCEFEKVCCKIPOVANGHVEVRKT--SNVQYQYVNIKCDKGRLOG 246
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 228 RGDVCTESGMRP-LPSCEEKSCDNPIYIPNGDYSPLRIKHTGDEITYOCRNFGFYPATRG 286
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 247 ETIPNMCKNGVWPALPTCEK-----PAPRGD-----MPHIDSGEDTST-----PSGRN 290
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 287 NFAKCTST 294
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 291 CNQNCCTS 298
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: November 21, 2000, 16:48:55  
 Job time: 333 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:43:48 ; Search time 49.92 Seconds  
(without alignments)  
293.168 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454  
Sequence: 1 EDCKGPPRENSSEILSGSMS.....DTYCTENGMSPPKCVRIK 428

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT:\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT:\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT:\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT:\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT:\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT:\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT:\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT:\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT:\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT:\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT:\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT:\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT:\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT:\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT:\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT:\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT:\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT:\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT:\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT:\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	40.3	240	18	W39154 Human partial Comp
2	918	37.4	216	18	W39155 Clione prrbp9H410 C
3	544.5	22.2	578	20	Y09065 Human complement f
4	481.5	19.6	581	12	R13490 Human C4 binding f
5	438	17.8	543	13	R28547 CRI-4 (52S, 53S, 5
6	437	17.8	453	20	Y55757 Human CRI protein
7	437	17.8	778	19	W73147 Amino acid sequenc
8	437	17.8	1930	19	W45899 Human complement r
9	437	17.8	2039	20	Y55751 Human C3b/C4b rece
10	437	17.8	2317	10	P92219 CRI protein. Homo
11	434	17.7	496	20	Y55752 Human CRI protein
12	434	17.7	543	13	R28557 CRI-4 (99H, 103E)

13	434	17.7	1537	12	R11982
14	434	17.7	2039	14	R36743
15	432	17.6	543	13	R28570
16	432	17.6	2039	12	R1810
17	431	17.6	450	20	Y55754
18	430	17.5	543	13	R28560
19	429	17.5	543	13	R28548
20	428	17.4	543	13	R28550
21	428	17.4	543	13	R28553
22	428	17.4	543	13	R28558
23	428	17.4	543	13	R28565
24	428	17.4	543	13	R28571
25	427	17.4	543	13	R28545
26	427	17.4	543	13	R28552
27	426	17.4	543	13	R28555
28	426	17.4	543	13	R28553
29	426	17.4	543	13	R28567
30	425	17.3	543	13	R28549
31	425	17.3	543	13	R28551
32	425	17.3	543	13	R28562
33	425	17.3	543	13	R28568
34	424	17.3	543	13	R28569
35	423	17.2	543	13	R28556
36	423	17.2	543	13	R28564
37	422	17.2	543	13	R28561
38	421	17.2	543	13	R28554
39	420	17.1	543	13	R28546
40	420	17.1	543	13	R28546
41	420	17.1	543	13	R28566
42	419	17.1	543	13	R28543
43	419	17.1	543	13	R28559
44	412	16.8	44	20	Y55756
45	409	16.7	450	20	Y55753

## ALIGNMENTS

RESULT 1	
ID W39154	standard; Protein; 240 AA.
AC W39154;	
XX	
DT 27-APR-1998	(first entry)
DE Human partial Complement factor H protein fragment 1.	
XX	
KW Complement factor H; tumour associated antigen; renal cancer; urogenital cancer; medicament; modulator.	
XX	
OS Homo sapiens.	
XX	
PN W09738136-A1.	
XX	
PD 16-OCT-1997.	
XX	
PF 09-APR-1997;	97WO-US05710.
XX	
PR 06-MAR-1997;	97US-0812481.
PR 09-APR-1996;	96US-0015083.
PR 09-APR-1996;	96US-0630048.
PR 06-MAR-1997;	97US-0038614.
PA (BARD-) BARD DIAGNOSTIC SCI INC.	
PI Enfield DL, Hass GM, Kinders RJ;	
XX	
DR WPI: 1997-512742/47.	
DR N-PSDB: V02790.	
XX	
PT Treating or screening for cancer, e.g. renal or urogenital cancer - by modulating or detecting tumour associated human complement factor	

PT H related antigen, or nucleic acid encoding it  
 XX  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein sequence represents a region of the human  
 CC tumour-associated complement factor H (CFH). This sequence is used  
 CC in the identification of complement factor H related proteins and  
 CC antigens isolated from clone pRRB9FH410 (see W39155). The detection of  
 CC such proteins and a CFH antigen can be used in screening or for the  
 CC treatment of renal or urogenital cancer, e.g. bladder, cervical or  
 CC prostate cancer. Agents that may modulate this antigen could be used in  
 CC the manufacture of a medicament for the treatment of a tumour cell.

SO Sequence 240 AA;

Query Match 40.3%; Score 990; DB 18; Length 240;  
 Best Local Similarity 71.7%; Pred. No. 3.2e-68;  
 Matches 172; Conservative 21; Mismatches 47; Indels 0; Gaps 0;

QY 78 FRLAVGSEFEFGAKVYITCDGEGQLLGEIDYRECDADGWTNDIPICEVVKCLPYLEENG 137  
 DB 1 fllggnvteygvkavylcneqy1lgelnrecedldgwtndipicevkc1pvtapeng 60  
 QY 138 RIVSGAEPDQEEYFGOVVRFECNSGFKIEGOKEMHCSEGLMSNKPQCVEISCLPPRY 197  
 DB 61 kvssamepdrfhyfvgavrfvcnsygllegdeemcsddgfwkpkcveiscspdv 120  
 QY 198 ENGDDGIYLYKPVYKERNRFOYKCKGQFVYKERDVACTGSGWNPQSCDEMTCLTPYIPNG 257  
 DB 121 lngsflsqkilykenerrfykcmgyesergdvclesgwrplpseeekscdnpylpug 180  
 QY 258 IYTPRIKIRIDETRECKNGFYPATRSPVSKCTTGTGIIAPRCGLKACDFQFQHGRI 317  
 DB 181 dypslrikrtgtdeltygcmrgfpatrgntactstgwlipaprc1kpcodypdikhg91 240

RESULT 2

ID W39155 standard; Protein; 216 AA.

AC W39155;

DT 27-APR-1998 (first entry)

DE Clone pRRB9FH410 CFH related protein fragment.

KW Complement factor H; tumour associated antigen; renal cancer;

KW urogenital cancer; medicament; modulator.

OS Synthetic.

PN W09738136-A1.

PD 16-OCT-1997.

PF 09-APR-1997; 97WO-US05710.

PR 06-MAR-1997; 97US-0812481.

PR 09-APR-1996; 96US-0015083.

PR 09-APR-1996; 96US-0630048.

PR 06-MAR-1997; 97US-0038614.

PA (BARD-) BARD DIAGNOSTIC SCI INC.

PI Enfield DL, Haas GM, Kinders RJ;

DR WPI: 1997-512742/47.

DR N-PSDB; W02791.

PT Treating or screening for cancer, e.g. renal or urogenital cancer -

PT by modulating or detecting tumour associated human complement Factor

PT H related antigen, or nucleic acid encoding it

XX  
 PS Example 6B; Fig 6B; 104pp; English.

CC This partial protein is found in clone pRRB9FH410 and represents a  
 CC complement factor H related protein with homology to a region of the  
 CC human tumour-associated complement factor H (CFH). The detection of this  
 CC protein and a CFH antigen can be used in screening or for the treatment  
 CC of renal or urogenital cancer, e.g. bladder, cervical or prostate cancer.  
 CC Agents that may modulate this antigen could be used in the manufacture of  
 CC a medicament for the treatment of a tumour cell.

SO Sequence 216 AA;

Query Match 37.4%; Score 918; DB 18; Length 216;  
 Best Local Similarity 73.1%; Pred. No. 8.8e-63;  
 Matches 158; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

QY 95 TCDEGYQLLGEIDYRECDADGWTNDIPICEVVKCLPYLEENGRIYSGAEPDQEEYFG 154  
 DB 1 tonegyq1lgelnrecedldgwtndipicevkc1pvtapengkvssamepdrfhyf 60  
 QY 155 VVRFECNSGFKIEGOKEMHCSEGLMSNKPQCVEISCLPPRYENGDIYLYKPVYKENER 214  
 DB 61 avrfvcnsygllegdeemcsddgfwkpkcveiscspdv1ngdysplrikrtgtdelty 120  
 QY 215 FOYKCKGQFVYKERDVACTGSGWNPQSCDEMTCLTPYIPNGIYTPRIKIRIDETIRY 274  
 DB 121 fgykcmgyesergdvclesgwrplpseeekscdnpylpugdysplrikrtgtdelty 180  
 QY 275 ECKNGFYPATRSPVSKCTTGTGIIAPRCGLKACDFP 310  
 DB 181 qerngfyatrgntactstgwlipaprc1kpcody 216

RESULT 3

ID Y09065 standard; Protein; 578 AA.

AC Y09065;

DT 06-JUL-1999 (first entry)

DE Human complement factor H homolog protein.

KW Human complement factor H; immunological mechanism; complement reaction;

KW gene therapy; immune stimulation; haematopoiesis regulation; chemotactic;

KW tissue growth activity; anti-inflammatory; tumour inhibition;

KW secretory signal.

OS Homo sapiens.

PN W09918200-A1.

PD 15-APR-1999.

PF 02-OCT-1998; 98WO-JP04448.

PR 06-OCT-1997; 97JP-0272837.

PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENT.

PI Kato S, Sekine S;

DR WPI: 1999-264019/22.

DR N-PSDB; X34737.

PT Human proteins with secretory signal sequences and nucleotide

PT sequences, useful in control of proliferation and differentiation of

PT cells

PS Claims 1; Page 55-58; 71pp; English.





Db 91 -fclykrctrrpplng--qveiktdlstrgsqlefscegflllgstlstr--ceqgdvgy 146  
 QY 116 WINDPICEVVKCLPTELENGRIVSGAEPDDEYVGVVFECSNGFKEGOKEMHCS 175  
 Db 147 wshpdpceivckkppddlngrf---hsgeenfyayfvtvscdprflilghaslst 202  
 QY 176 -EN---GIMSNEKPOCEVISCLEPRVENGGIT-LKPVYKENERFOYKCKOGFYVERGD 230  
 Db 203 venetlgvwrpspplceklctcrkpdvshgemvsgfiplynkdtlvcfkcgkgtv1--rgs 260  
 QY 231 AV--C-TGSQWNPQ-PSCEMTCLT-PIYNGIYT--PHRIK---HRIDEIRYCKNGF 280  
 Db 261 svlhcdadckwmpspaccpncscinlpdiphaswetyprptkedvyyvgtvltvtrcmgy 320  
 QY 281 YRATKSPVSKCTITG--WIPAPRCSLKPCDFQPKHGRLYEESRRP----- 325  
 Db 321 kpltdptlvicqknlrvtprygccealcepeklmgngeltqhrktrpanhcvfygdeis 380  
 QY 326 -----YRPVPIG-----KEYSYCDN 341  
 Db 381 fchetsrfsalcqgdqgtvsptpscgdicnfpkiahghyqssysffikeeltyeckd 440  
 QY 342 GFTTSPQSYWDYLRCTVNGMEBEVP-CLRQCTFHVEYEGSSYWRRTYEGOSAKVQCHS 400  
 Db 441 gyllyvga---klscsyshwsapapqckalcrkpelvnglrvldkqgyvpenvtlqcds 497  
 QY 401 GYSLRNGDITYCTEN-GMSPP-PKC 424  
 Db 498 gygvv-qpqslcsqurtwlypevpkc 522

RESULT 5  
 R28547  
 ID R28547 standard; peptide: 543 AA.  
 AC R28547;  
 DT 19-MAR-1993 (first entry)  
 DE CRI-4 (52S, 53S, 54P) analogue.  
 KW short consensus repeat; regulator of complement activation;  
 KW C3b binding; C4b binding; human complement type 1 receptor.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT 1..60 /label= SCR-1  
 FT 61..122 /label= SCR-2  
 FT 451..510 /label= SCR-8  
 FT 511..543 /label= SCR-9  
 FT /label= SCR-9  
 FT Misc-difference 52 /note= "TRUNCATED"  
 FT Misc-difference 53 /note= "Thr substituted by Ser (SCR-8)"  
 FT Misc-difference 54 /note= "Gly substituted by Ser (SCR-8)"  
 FT Misc-difference 54 /note= "Ala substituted by Pro (SCR-8)"  
 XX EP512733-A.  
 XX 11-NOV-1992.  
 XX 28-APR-1992; 92EP-0303826.  
 XX 03-MAY-1991; 91US-0695514.  
 XX (UNIM ) UNIV WASHINGTON.

XX PI Atkinson JP, Hourcade D, Krych M;  
 XX DR WPI; 1992-375009/46.  
 PT Complement activity regulator protein analogues - useful for  
 PT treating auto-immune diseases, to suppress transplant rejection,  
 PT for diagnosis etc.  
 PS Claim 11; Fig 2 and R11810; 23pp; English.  
 CC The CDNA clone designated CRI-4 was described in J. Exp. Med. (1988)  
 CC 168:1255-1270. It encodes the first 8 and a half amino terminal  
 CC SCRs of CRI. The invention concerns analogues of "regulator of  
 CC complement activation" proteins or truncated, hybrid or recombinant  
 CC forms of them. CRI-4 is a preferred truncated form and a number of  
 CC specified substitution variants of it are claimed in which certain  
 CC positions in SCR-1 which have been identified as important for the  
 CC degree of C3b- and C4b-binding are substituted by amino acids from  
 CC the corresponding positions in SCR-8. The specification does not  
 CC contain the CRI-4 sequence; the sequence given here was constructed  
 CC from the full-length CRI amino acid sequence having GENSESO  
 CC accession number R11810 and descriptions in the disclosure.  
 XX  
 SQ Sequence 543 AA;

Query Match 17.8%; Score 438; DB 13; Length 543;  
 Best local similarity 27.9%; Pred. No. 1,le-25;  
 Matches 133; Conservative 63; Mismatches 146; Indels 132; Gaps 31;

QY 27 GTQATYKCRPGRTKGTIVKCKNGEWPNSRLCRKRPCHGPTPGSPRLAVSEF 86  
 Db 25 gtylmyecrpygsgrpfslclksnw--sspkdrctrcrknpprvngmwhvlykg--1 80  
 QY 87 EFGAKVVTCTDGGYOLGIEDYRECDAGS---WINDPICEVVK-LPVTLENGRIYS 141  
 Db 81 qfsgqtkysctgyrligs-satclisgdtvawdetrpdrtpcglpdt-iltgdtis 138  
 QY 142 GAEPDQEVYFGQVNFECNSG-----FKIEGOKEMHCSEN---GLMSNEKPOC-VBI 190  
 Db 139 tnre--nfhygsvtyrcnpgsgarkvfelvgepsiyctsnddqyivsgpappclpn 195  
 QY 191 SCLPRVENG---DGIVLKPYKENERQYKCKOGFYVKKEGDAVCTG-SCGMNQ-PSC 244  
 Db 196 kctppvengllvsdn---tslfsinevetrcqpfymkprvkcqalnkwepepisc 252  
 QY 245 EEMTCLTYIPNGIYTPHRIKHRIID---DEIRYCKNGFYPATRSPVSKCTITG-WIP- 298  
 Db 253 svr-cqpp--pdvlhaertqrkdnlfpqgevfyccepg-ydlirgaamrctprgdwspa 308  
 QY 299 ARCSLKPCD--FPOFKHGRLYEESRRPFPV--PICKEYSYVCDNGFTT--PSQSYWD 352  
 Db 309 pftcevkscddfmqllngvr-----lfpvnlqlyakvdfvcddegfqlkysasy-- 358  
 QY 353 YLRCTVNG-----WEPEVPCLRQCIFHYVEYEGSSYWRRTYEGOSAKVQCHGYSLPN-- 406  
 Db 359 ---cvlagmeslmsvpyceq-----ifcspvvlmgvr 390  
 QY 407 -----GDDTYVCTE-----NG-W-SPPPKC 424  
 Db 391 htgkplevfpfgkavnycdphdrtgtsfdllgestlctsdapngnywsspaprc 446

RESULT 6  
 ID Y55757 standard; Protein; 453 AA.  
 AC Y55757;  
 XX 22-FEB-2000 (first entry)  
 XX Human CRI protein LHR-C fragment.

XX C3B/C4b receptor; CRI protein; cell-surface protein; erythrocyte; human;  
 KW complement regulatory activity; complement pathway enzyme; tissue damage;  
 KW reperfusion injury; Arthus reaction; myocardial infarct; inflammation;  
 KW heat condition; autoimmune disorder; long homologous repeat; LHR.  
 OS Homo sapiens.  
 XX US5981481-A.  
 XX  
 PD 09-NOV-1999.  
 XX  
 PF 06-JUN-1995; 95US-0470652.  
 XX  
 PR 03-APR-1989; 89US-0332865.  
 PR 06-DEC-1974; 74US-0350238.  
 PR 24-FEB-1993; 93US-0026134.  
 PR 01-APR-1988; 88US-0176532.  
 XX  
 PA (UYJO ) UNIV JOHNS HOPKINS.  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.  
 XX  
 PI Conclino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, Ip SH;  
 PI Marsh HC, Carson GR;  
 XX  
 DR WPI: 1999-633357/54.  
 XX  
 PT A human C3B/C4b receptor (CRI) protein having antiinflammatory and  
 PT cardiant activity -  
 XX  
 PS Disclosure: Fig 5B; 87pp; English.  
 XX  
 CC The invention relates to a human C3B/C4b receptor (CRI) protein. The CRI  
 CC protein or fragment is expressed as a cell-surface protein on the surface  
 CC of a non-human cell and exhibits a complement regulatory activity of full  
 CC -length human CRI as expressed on erythrocytes. The CRI function in vivo  
 CC may be mediated through the inhibition of complement pathway enzymes. The  
 CC soluble CRI protein exhibits a complement regulatory activity, and this  
 CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and  
 CC neutrophil mediated tissue damage, and reduce myocardial infarct size,  
 CC and inflammation. The CRI protein and its fragments can also be used in  
 CC the treatment of conditions which involve unwanted complement activity,  
 CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,  
 CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti-  
 CC -CRI antibodies are used in assays, and diagnostics. The present sequence  
 CC represents the human CRI protein long homologous repeat (LHR)-C fragment.  
 XX  
 SQ Sequence 453 AA;  
 Query Match 17.8%; Score 437; DB 20; Length 453;  
 Best Local Similarity 29.6%; Pred. No. 1e-25;  
 Matches 136; Conservative 64; Mismatches 161; Indels 98; Gaps 32;

DB 254 srvcqppellhgehtpsndfnspgqevfysocp9-ydlrsgaalslctpqgdwsppear 312  
 QY 302 CSEKPCD-EPQKRGRLYEESRRPYF--VPICKYSYCDNGFTTPSQSYWDYLRCT 357  
 DB 313 cavsccddflgqlphgfv-----lfplnlqgkavsfvcdgfrlkgs--vshcv 361  
 QY 358 VNG---WEPEVCLRCICF-----HY-----VEYGSYSWQRRYLEGQSAKV 396  
 DB 362 lvgfhrsImnsvpvcch-llcpnpallngfthgtpsgdldpgeIsny----- 408  
 QY 397 QC--HSGYSLP---NGDFTYCTE---NG-W-SPPPKC 424  
 DB 409 tcdphdrgmtfnlgestlrcstsdphngnywswspaprc 447

RESULT 7  
 W73147  
 ID W73147 standard; protein; 778 AA.  
 XX  
 AC W73147;  
 XX  
 DT 29-JAN-1999 (first entry)  
 XX

DE Amino acid sequence of the soluble complement receptor 1 (sCR1).  
 XX

XX Human: soluble complement receptor 1; sCR1; T-cell; B-cell;  
 KW mediated immune response; inhibition; tissue rejection; gene therapy;  
 KW dystrophin; inflammatory response; interferon gamma secretory response;  
 KW autoimmune response; neurological response; Alzheimer's disease;  
 KW Parkinson's disease; multiple sclerosis; systemic lupus erythematosus;  
 KW rheumatoid arthritis; myasthenia gravis; epidermis bullosa;  
 KW Hashimoto's disease.  
 XX  
 OS Homo sapiens.  
 OS  
 XX MO9845430-A1.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 06-APR-1998; 98WO-GB01012.  
 XX  
 PR 05-APR-1997; 97GB-0006950.  
 XX  
 PA (ANNE/) ANNENKOV A.  
 PA (CHER/) CHERNAJOVSKY Y.  
 XX  
 PI Annenkov A, Chernajovsky Y;  
 XX  
 DR WPI: 1998-568350/48.  
 XX

PT Fragment of soluble human complement receptor 1 - useful for  
 PT treating T-cell or B-cell mediated immune responses e.g.  
 PT inflammatory responses such as rheumatoid arthritis  
 XX  
 PS Disclosure: Fig 1; 54pp; English.  
 XX

CC This is an amino acid sequence of the human soluble complement  
 CC receptor 1 (sCR1), useful in the treatment of T-cell or B-cell  
 CC mediated immune responses. It is used to inhibit a T-cell or  
 CC B-cell-mediated immune response to prevent immune response-mediated  
 CC tissue rejection and destruction or clearance or inactivation of an  
 CC expressed protein especially from cells that have been treated by gene  
 CC therapy to express the protein, e.g. dystrophin. The protein can also  
 CC be used to inhibit a T-cell or B-cell-mediated inflammatory response,  
 CC an interferon-gamma secretory response, autoimmune response or  
 CC neurological response, e.g. Alzheimer's or Parkinson's disease or  
 CC multiple sclerosis. Also the protein can be used to treat systemic  
 CC lupus erythematosus, rheumatoid arthritis, myasthenia gravis,  
 CC epidermis bullosa or Hashimoto's disease.  
 CC  
 SQ Sequence 778 AA;





```
Db 75 gtylnyecrpgysgrfslclksnwtgkdr--crkscsrppdpvngmvhikg--i 130
OY 87 ERFKAVVYTTDEGQOLLGEIDYRECDAG---WTNDIPICEVVKC-LPYTELENGRIYS 141
Db 131 qfsgsklyscckgyrlrlgs--ssatclisgdtvldnepricdrpcglppt-ltngdfis 188
OY 142 GAAPDDEYFGOVVREFECNSG-----FKIEGOKEMHCSEN-----GLMSNEKPOC-VEI 190
Db 189 tne-----nfhygsavvtlycnpysggrkvfelvgepslyctsnddqyglwspapqcllpn 245
OY 191 SCLPRPVENG---DGITYLKPVYKENERFQYKCKOGFYVKEKGDVACTG--SGWNPQ--PSC 244
Db 246 kcpvnevgilvsdn---zslfslnevefrcqpfvmkgrprvkqgalnkewelpsc 302
OY 245 EEMTCLPYIPNGIYTPHRIKHKRID----DEIRYCKNGFYPATRSVSKCTITG-WIP- 298
Db 303 srvcqpp--pdvlnhaertgrdknfspgqevlyscpg-ydlrgaasmrctpgqdwspa 358
OY 299 APRCSLKPCD--PPOFKHGRLYEESRRPYPV--PIGKEYSYCDNGFTT--PSQSYMD 352
Db 359 apccevksccddfmngqllntrv-----lfvnlqlgakvdtvcdqegtlkgsasay-- 408
OY 353 YLRCTVNG---WEPEVPCLRQCIFHYVEYGESSYWMQRYIEGOSAKVQCHSGYSLPN-- 406
Db 409 ---cvlagmeslwnssvpcq-----lfcpvprvlpngr 440
OY 407 -----GODTYTCTE---NG-W-SPPKPC 424
Db 441 hngkplevrfpfgkavnytcphdpdrgtsfdligestlrcsdpgngvwsppaprc 496
RESULT 11
Y55752
ID Y55752 standard; Protein: 496 AA.
AC Y55752;
DE 22-FEB-2000 (first entry)
XX Human CRI protein LHR-A SCR fragment.
XX C3B/C4b receptor; CRI protein; cell-surface protein; erythrocyte; human;
XX complement regulatory activity; complement pathway enzyme; tissue damage;
XX reperfusion injury; Arthus reaction; myocardial infarct; inflammation;
XX heart condition; autoimmune disorder; long homologous repeat; LHR; SCR;
XX short consensus repeat.
XX OS Homo sapiens.
XX PM US5981481-A.
XX PD 09-NOV-1999.
XX PE 06-JUN-1995; 95US-0470652.
XX PR 03-APR-1989; 89US-0332865.
XX PR 06-DEC-1974; 74US-0350238.
XX PR 24-FEB-1993; 93US-0026134.
XX PR 01-APR-1988; 88US-0176532.
XX PA (UYJO ) UNIV JOHNS HOPKINS.
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX PA (AVAN-) AVANT IMMUNOTHERAPEUTICS INC.
XX PI Conclino MF, Wong WM, Makrides SC, Klickstein LB, Fearon DT, IP SH;
XX PI Marsh HC, Carson GR;
XX DR WPI; 1999-633357/54.
XX A human C3B/C4b receptor (CRI) protein having antiinflammatory and
XX PT cardiant activity -
XX OS Homo sapiens.
```

```
PS Disclosure; Fig 10; 87pp; English.
XX The invention relates to a human C3B/C4b receptor (CRI) protein. The CRI
CC protein or fragment is expressed as a cell-surface protein on the surface
CC of a non-human cell and exhibits a complement regulatory activity of full
CC -length human CRI as expressed on erythrocytes. The CRI function in vivo
CC may be mediated through the inhibition of complement pathway enzymes. The
CC soluble CRI protein exhibits a complement regulatory activity, and this
CC may be used to prevent reperfusion injury, inhibit Arthus reaction, and
CC neutrophil mediated tissue damage, and reduce myocardial infarct size,
CC and inflammation. The CRI protein and its fragments can also be used in
CC the treatment of conditions which involve unwanted complement activity,
CC e.g. shock lung, tissue damage due to burn, or ischemic heart conditions,
CC and autoimmune disorders. CRI proteins, analogues, derivatives, and anti-
CC -CRI antibodies are used in assays, and diagnostics. The present sequence
CC represents the short consensus repeat (SCR) fragments of human CRI
XX protein long homologous repeat (LHR)-A sequence.
SQ Sequence 496 AA:
Query Match 17.7%; Score 434; DB 20; Length 496;
Best Local Similarity 27.8%; Pred. No. 2e-25;
Matches 133; Conservative 63; Mismatches 148; Indels 134; Gaps 32;
OY 27 GTQATYKCRPGYRTLTGTYVVC-KNGEWPNSPRICRKRPGDTPGCSFRLAVGSE 85
Db 71 gtylnyecrpgysgrfslclksnwtgkdr--crkscsrppdpvngmvhikg-- 126
OY 86 FEFKAVVYTTDEGQOLLGEIDYRECDAG---WTNDIPICEVVKC-LPYTELENGRIY 140
Db 127 idfgsqklyscckgyrlrlgs--ssatclisgdtvldnepricdrpcglppt-ltngdfi 184
OY 141 SGAAPDDEYFGOVVREFECNSG-----FKIEGOKEMHCSEN-----GLMSNEKPOC-VEI 189
Db 185 stnre---nfhygsavvtlycnpysggrkvfelvgepslyctsnddqyglwspapqcllp 241
OY 190 ISCLPRPVENG---DGITYLKPVYKENERFQYKCKOGFYVKEKGDVACTG--SGWNPQ--PS 243
Db 242 nktcprpvnevgilvsdn---zslfslnevefrcqpfvmkgrprvkqgalnkewelpsp 298
OY 244 CEEMTCLPYIPNGIYTPHRIKHKRID----DEIRYCKNGFYPATRSVSKCTITG-WIP 298
Db 299 srvcqpp--pdvlnhaertgrdknfspgqevlyscpg-ydlrgaasmrctpgqdwspa 354
OY 299 -APRCSLKPCD--PPOFKHGRLYEESRRPYPV--PIGKEYSYCDNGFTT--PSQSY 350
Db 355 aapccevksccddfmngqllntrv-----lfvnlqlgakvdtvcdqegtlkgsasay 406
OY 351 WDYLRCCTVNG---WEPEVPCLRQCIFHYVEYGESSYWMQRYIEGOSAKVQCHSGYSLPN 406
Db 407 -----cvlagmeslwnssvpcq-----lfcpvprvlpn 436
OY 407 -----GODTYTCTE---NG-W-SPPKPC 424
Db 437 grhtgkplevrfpfgkavnytcphdpdrgtsfdligestlrcsdpgngvwsppaprc 494
RESULT 12
R28557
ID R28557 standard; peptide; 543 AA.
AC R28557;
XX 19-MAR-1993 (first entry)
XX CRI-4 (99H, 103E) analogue.
XX Short consensus repeat; regulator of complement activation;
XX C3b binding; C4b binding; human complement type 1 receptor.
XX OS Homo sapiens.
```

[illegible]

Dd	253	srv-cgpp--pdvIhaertqrkdnafpqgevfyscepg-ydlrگاasmrctpggdwspa	308
Oy	299	APRCSLAPCD--PFOFKHGRGLYTEESRRPYFPV--PIGKEYSYICDNGFTT--PSOSYMD	352
Dd	309	aptcevksccddfmqllngvr-----lfpvnlqlgakvdrvcdeglfqlkgsasasy--	358
Oy	353	YLRCRYNS----WPEPVCLNQCIHFHYVEGESSYWQRRIEGSAAKVOCHSGVSLPN--	406
Dd	359	--cvlaqmeslwnssvpvceq-----ifcpsspvlpngr	390
Oy	407	-----GDDTYCTE---NG-W-SPPPKC	424
Dd	391	htgkplevfipgkavnyrcdpbdrgtsfdllgestllrctsdppqngvwspaprc	446
RESULT	13		
R11982	ID	R11982 standard; Protein; 1537 AA.	
XX	AC	R11982;	
XX	Dt	25-JUN-1991 (first entry)	
XX	DE	Partial human complement type 1 receptor.	
XX	KM	Complement system; C3b/C4b receptor; CRI; allergic reaction;	
XX	KW	Immune response; Long homologous repeat; LHR.	
XX	OS	Homo sapiens.	
XX	FH	Key	Location/Qualifiers
FT	Region	1..438	/label= LHR-B
FT	Region	439..891	/label= LHR-C
FT	Region	892..1341	/label= LHR-D
FT	Region	1495..1498	/note= "positively-charged; preceded by hydrophobic sequence"
FT	Region	1521..1526	/note= "has 67 per cent homology to site of protein kinase C phosphorylation in the EGF receptor"
XX	PD	WO9105047-A.	
XX	PD	18-APR-1991.	
XX	PF	25-SEP-1990; 90MO-US05454.	
XX	PR	26-SEP-1989; 89US-0412745.	
XX	PR	26-SEP-1990; 90US-0912349.	
PA	(TCBL-) T CELL SCI INC.		
PA	(UYUO ) JOHNS HOPKINS UNIVERSITY.		
PA	(BRIG-) BRIGHAM AND WOMEN'S HOSPITAL.		
PI	Featon DT, Kluckstein LB, Wong WM, Carson GR, Hoh M, Concino MF		
PI	Makrides SC, Marsh HC;		
DR	N-PSDB; Q11643.		
XX	WPI; 1991-132854/18.		
PT	Human complement receptor type 1 gene, encoded proteins and fragments - for treatment of immune disorders, myocardial infarct,		
XX	damage due to inflammation and in treatment of thrombosis		
XX	Disclosure; Fig 5; 234p; English.		
CC	This sequence comprises three of the four tandem, direct, long		
CC	homologous repeats of the full-length F allozyme of CRI. LHR-A is		
CC	absent. Each LHR might represent a single C3b/C4b binding domain,		

THE UNIVERSITY OF CHICAGO

966 gtslkya

966 qtsj kvecrnawarfa{tc}dn] wu=-cckdrcbkckctccdwurc auh:4.3! 1.001





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:40 ; Search time 47.7 Seconds

(without alignments)  
569,412 Million cell updates/sec

Title: US-09-316-163-14

2454

Sequence: 1 ECKCKGPPRENSSEILSGSWS.....DIIYCTENGWSPPKCVRIK 428

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.65:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2055	83.7	1234	1	NBM5H	complement factor
2	1714	69.8	1231	1	NBM5H	complement factor
3	1709	69.6	449	1	NBM5H	complement factor
4	1242	50.6	669	2	S65551	factor H - bovine
5	813	33.1	452	2	A35068	complement factor
6	811	33.0	303	2	H35068	apolipoprotein H-r
7	569.5	23.2	1053	2	S46199	probable complement
8	566	23.1	808	2	D35069	complement factor
9	529.5	21.6	597	1	S53711	C4BP alpha chain P
10	481.5	19.6	597	1	NBM5H	C4BP binding protei
11	474.5	19.3	610	1	I46001	C4b-binding protei
12	466.5	19.0	331	2	A45222	complement factor
13	453	18.5	676	2	A45900	complement C3b rec
14	449	18.3	558	2	S57953	C4BP protein alpha
15	447	18.2	2014	2	I36936	complement recepto
16	437	17.8	2489	2	I73012	complement C3b/C4b
17	409	16.7	497	2	JC2054	complement regulat
18	400.5	16.3	482	2	A34924	complement C3b/C4b
19	396	16.1	579	2	A56740	sperm-egg recogniti
20	384.5	15.7	469	1	NBM5H	C4b-binding protei
21	375	15.3	560	2	T16833	hypothetical prote
22	373	15.2	1025	1	A43526	complement C3d/Bps
23	363	14.8	1091	1	PI0009	complement C3d/Bps
24	336	13.7	661	1	KFH013	coagulation factor
25	332	13.5	668	2	A46013	coagulation factor
26	321	13.1	433	2	A30550	complement C3b/C4b
27	319	13.0	443	2	G35070	apolipoprotein H-r
28	318	13.0	440	2	A43519	complement recepto
29	315	12.8	330	2	I56100	complement factor

30	314.5	12.8	263	1	WMV2SP	apolipoprotein H h
31	310.5	12.7	768	2	A42755	P-selectin precurs
32	310	12.6	270	2	I37278	complement factor
33	309.5	12.6	830	2	A30359	P-selectin precurs
34	298.5	12.2	377	2	I54479	membrane cofactor
35	298.5	12.2	384	2	S01896	membrane cofactor
36	297.5	12.1	263	1	C36838	complement control
37	297.5	12.1	263	2	T28450	hypothetical prote
38	296.5	12.1	263	2	B72152	B18L protein - var
39	296	12.1	768	2	I53821	P-selectin - rat
40	294.5	12.0	362	2	JC5194	membrane cofactor
41	294.5	12.0	369	2	JC5138	membrane cofactor
42	293.5	12.0	369	2	G02913	sperm CD46 - human
43	293.5	12.0	369	2	I57998	membrane cofactor
44	290.5	11.8	360	2	T42921	complement control
45	290	11.8	610	2	A35046	E-selectin precurs

#### ALIGNMENTS

RESULT 1

NBM5H

complement factor H precursor - mouse

N:Alternate names: protein beta-1-H

C:Species: Mus musculus (house mouse)

C>Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 22-Jun-1999

C:Accession: A26154; I49711; I49728

R:Kristensen, T.; Tack, B.F.

Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986

A>Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length

A:Reference number: A26154; MUID:8623353

A:Accession: A26154

A:Molecule type: mRNA

A:Residues: 1-1234 <R1>

A:Cross-references: GB:M12660; NID:G193724; PIDN:AAA37759.1; PID:G387181

R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.

J. Immunol. 144, 358-362, 1990

A>Title: Demonstration of an unusual allelic variation of mouse factor H by the compl

A:Reference number: I49711; MUID:90111033

A:Accession: I49711

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-18 <RES>

A:Cross-references: GB:M31979; NID:G193726; PIDN:AAA37762.1; PID:G193729

R:Munoz-Canoaves, P.; Tack, B.F.; Viik, D.P.

Biochemistry 28, 9891-9897, 1989

A>Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma

A:Reference number: I49728; MUID:90148935

A:Accession: I49728

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-19 <RE2>

A:Cross-references: GB:J02891; NID:G193805; PIDN:AAA37795.1; PID:G553926

C:Comment: Two codominant alleles of factor H are present in mice.

C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot

(C5 convertase) in the alternative complement pathway.

C:Genetics:

A:Map position: 1

C:Superfamily: complement factor H; complement factor H repeat homology

C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1234/Product: complement factor H #status predicted <MP>

F:21-80/Domain: complement factor H repeat homology <FH01>

F:85-144/Domain: complement factor H repeat homology <FH02>

F:146-205/Domain: complement factor H repeat homology <FH03>

F:210-262/Domain: complement factor H repeat homology <FH04>

F:246-248/Region: cell attachment (R-G-D) motif

F:325-385/Domain: complement factor H repeat homology <FH05>

F:389-442/Domain: complement factor H repeat homology <FH06>

F:448-505/Domain: complement factor H repeat homology <FH08>

F:509-564/Domain: complement factor H repeat homology <FH09>

F:569-622/Domain: complement factor H repeat homology <FH10>  
 F:629-683/Domain: complement factor H repeat homology <FH11>  
 F:690-743/Domain: complement factor H repeat homology <FH12>  
 F:752-802/Domain: complement factor H repeat homology <FH13>  
 F:808-861/Domain: complement factor H repeat homology <FH14>  
 F:867-931/Domain: complement factor H repeat homology <FH15>  
 F:936-989/Domain: complement factor H repeat homology <FH16>  
 F:994-1048/Domain: complement factor H repeat homology <FH17>  
 F:1053-1107/Domain: complement factor H repeat homology <FH18>  
 F:1114-1168/Domain: complement factor H repeat homology <FH19>  
 F:1172-1233/Domain: complement factor H repeat homology <FH20>  
 F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357-402,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1082-1107,1114-1168,1172-1233,1237-1262,1267-1309,1314-1368,1373-1427,1432-1486,1491-1545,1550-1604,1609-1663,1668-1722,1727-1781,1786-1840,1845-1899,1904-1958,1963-2017,2022-2076,2081-2135,2140-2194,2199-2253,2258-2312,2317-2371,2376-2430,2435-2489,2494-2548,2553-2607,2612-2666,2671-2725,2730-2784,2789-2843,2848-2902,2907-2961,2966-3020,3025-3079,3084-3138,3143-3197,3202-3256,3261-3315,3320-3374,3379-3433,3438-3492,3497-3551,3556-3610,3615-3669,3674-3728,3733-3787,3792-3846,3851-3905,3910-3964,3969-4023,4028-4082,4087-4141,4146-4200,4205-4259,4264-4318,4323-4377,4382-4436,4441-4495,4500-4554,4559-4613,4618-4672,4677-4731,4736-4790,4795-4849,4854-4908,4913-4967,4972-5026,5031-5085,5090-5144,5149-5203,5208-5262,5267-5321,5326-5380,5385-5439,5444-5498,5503-5557,5562-5616,5621-5675,5680-5734,5739-5793,5798-5852,5857-5911,5916-5970,5975-6029,6034-6088,6093-6147,6152-6206,6211-6265,6270-6324,6329-6383,6388-6442,6447-6501,6506-6560,6565-6619,6624-6678,6683-6737,6742-6796,6801-6855,6860-6914,6919-6973,6978-7032,7037-7091,7096-7150,7155-7209,7214-7268,7273-7327,7332-7386,7391-7445,7450-7504,7509-7563,7568-7622,7627-7681,7686-7740,7745-7799,7804-7858,7863-7917,7922-7976,7981-8035,8040-8094,8099-8153,8158-8212,8217-8271,8276-8330,8335-8389,8394-8448,8453-8507,8512-8566,8571-8625,8630-8684,8689-8743,8748-8802,8807-8861,8866-8920,8925-8979,8984-9038,9043-9097,9102-9156,9161-9215,9220-9274,9279-9333,9338-9392,9397-9451,9456-9510,9515-9569,9574-9628,9633-9687,9692-9746,9751-9805,9810-9864,9869-9923,9928-9982,9987-10041,10046-10100,10105-10159,10164-10218,10223-10277,10282-10336,10341-10395,10400-10454,10459-10513,10518-10572,10577-10631,10636-10690,10695-10749,10754-10808,10813-10867,10872-10926,10931-10985,10990-11044,11049-11103,11108-11162,11167-11221,11226-11280,11285-11339,11344-11398,11403-11457,11462-11516,11521-11575,11580-11634,11639-11693,11698-11752,11757-11811,11816-11870,11875-11929,11934-11988,11993-12047,12052-12106,12111-12165,12170-12224,12229-12283,12288-12342,12347-12401,12406-12460,12465-12519,12524-12578,12583-12637,12642-12696,12701-12755,12760-12814,12819-12873,12878-12932,12937-12991,12996-13050,13055-13109,13114-13168,13173-13227,13232-13286,13291-13345,13350-13404,13409-13463,13468-13522,13527-13581,13586-13640,13645-13699,13704-13758,13763-13817,13822-13876,13881-13935,13940-13994,13999-14053,14058-14112,14117-14171,14176-14230,14235-14289,14294-14348,14353-14407,14412-14466,14471-14525,14530-14584,14589-14643,14648-14702,14707-14761,14766-14820,14825-14879,14884-14938,14943-14997,15002-15056,15061-15115,15120-15174,15179-15233,15238-15292,15297-15351,15356-15410,15415-15469,15474-15528,15533-15587,15592-15646,15651-15705,15710-15764,15769-15823,15828-15882,15887-15941,15946-16000,16005-16059,16064-16118,16123-16177,16182-16236,16241-16295,16300-16354,16359-16413,16418-16472,16477-16531,16536-16590,16595-16649,16654-16708,16713-16767,16772-16826,16831-16885,16890-16944,16949-16993,16998-17052,17057-17111,17116-17170,17175-17229,17234-17288,17293-17347,17352-17406,17411-17465,17470-17524,17529-17583,17588-17642,17647-17701,17706-17760,17765-17819,17824-17878,17883-17937,17942-17996,18001-18055,18060-18114,18119-18173,18178-18232,18237-18291,18296-18350,18355-18409,18414-18468,18473-18527,18532-18586,18591-18645,18650-18704,18709-18763,18768-18822,18827-18881,18886-18940,18945-18999,19004-19058,19063-19117,19122-19176,19181-19235,19240-19294,19299-19353,19358-19412,19417-19471,19476-19530,19535-19589,19594-19648,19653-19707,19712-19766,19771-19825,19830-19884,19889-19943,19948-19992,19997-20051,20056-20110,20115-20169,20174-20228,20233-20287,20292-20346,20351-20405,20410-20464,20469-20523,20528-20582,20587-20641,20646-20700,20705-20759,20764-20818,20823-20877,20882-20936,20941-20995,21000-21054,21059-21113,21118-21172,21177-21231,21236-21290,21295-21349,21354-21408,21413-21467,21472-21526,21531-21585,21590-21644,21649-21703,21708-21762,21767-21821,21826-21880,21885-21939,21944-21998,22003-22057,22062-22116,22121-22175,22180-22234,22239-22293,22298-22352,22357-22411,22416-22470,22475-22529,22534-22588,22593-22647,22652-22706,22711-22765,22770-22824,22829-22883,22888-22942,22947-22991,22996-23050,23055-23109,23114-23168,23173-23227,23232-23286,23291-23345,23350-23404,23409-23463,23468-23522,23527-23581,23586-23640,23645-23699,23704-23758,23763-23817,23822-23876,23881-23935,23940-23994,23999-24053,24058-24112,24117-24171,24176-24230,24235-24289,24294-24348,24353-24407,24412-24466,24471-24525,24530-24584,24589-24643,24648-24702,24707-24761,24766-24820,24825-24879,24884-24938,24943-24997,24998-25052,25057-25111,25116-25170,25175-25229,25234-25288,25293-25347,25352-25406,25411-25465,25470-25524,25529-25583,25588-25642,25647-25701,25706-25760,25765-25819,25824-25878,25883-25937,25942-25996,26001-26055,26060-26114,26119-26173,26178-26232,26237-26291,26296-26350,26355-26409,26414-26468,26473-26527,26532-26586,26591-26645,26650-26704,26709-26763,26768-26822,26827-26881,26886-26940,26945-26999,27004-27058,27063-27117,27122-27176,27181-27235,27240-27294,27299-27353,27358-27412,27417-27471,27476-27530,27535-27589,27594-27648,27653-27707,27712-27766,27771-27825,27830-27884,27889-27943,27948-27992,27997-28051,28056-28110,28115-28169,28174-28228,28233-28287,28292-28346,28351-28405,28410-28464,28469-28523,28528-28582,28587-28641,28646-28700,28705-28759,28764-28818,28823-28877,28882-28936,28941-28995,29000-29054,29059-29113,29118-29172,29177-29231,29236-29290,29295-29349,29354-29408,29413-29467,29472-29526,29531-29585,29590-29644,29649-29703,29708-29762,29767-29821,29826-29880,29885-29939,29944-29998,30003-30057,30062-30116,30121-30175,30180-30234,30239-30293,30298-30352,30357-30411,30416-30470,30475-30529,30534-30588,30593-30647,30652-30706,30711-30765,30770-30824,30829-30883,30888-30942,30947-30991,30996-31050,31055-31109,31114-31168,31173-31227,31232-31286,31291-31345,31350-31404,31409-31463,31468-31522,31527-31581,31586-31640,31645-31699,31704-31758,31763-31817,31822-31876,31881-31935,31940-31994,31999-32053,32058-32112,32117-32171,32176-32230,32235-32289,32294-32348,32353-32407,32412-32466,32471-32525,32530-32584,32589-32643,32648-32702,32707-32761,32766-32820,32825-32879,32884-32938,32943-32997,33002-33056,33061-33115,33120-33174,33179-33233,33238-33292,33297-33351,33356-33410,33415-33469,33474-33528,33533-33587,33592-33646,33651-33705,33710-33764,33769-33823,33828-33882,33887-33941,33946-33990,33995-34049,34054-34108,34113-34167,34172-34226,34231-34285,34290-34344,34349-34403,34408-34462,34467-34521,34526-34580,34585-34639,34644-34698,34703-34757,34762-34816,34821-34875,34880-34934,34939-34993,35000-35054,35059-35113,35118-35172,35177-35231,35236-35290,35295-35349,35354-35408,35413-35467,35472-35526,35531-35585,35590-35644,35649-35703,35708-35762,35767-35821,35826-35880,35885-35939,35944-35998,36003-36057,36062-36116,36121-36175,36180-36234,36239-36293,36298-36352,36357-36411,36416-36470,36475-36529,36534-36588,36593-36647,36652-36706,36711-36765,36770-36824,36829-36883,36888-36942,36947-36991,36996-37050,37055-37109,37114-37168,37173-37227,37232-37286,37291-37345,37350-37404,37409-37463,37468-37522,37527-37581,37586-37640,37645-37699,37704-37758,37763-37817,37822-37876,37881-37935,37940-37994,37999-38053,38058-38112,38117-38171,38176-38230,38235-38289,38294-38348,38353-38407,38412-38466,38471-38525,38530-38584,38589-38643,38648-38702,38707-38761,38766-38820,38825-38879,38884-38938,38943-38997,39002-39056,39061-39115,39120-39174,39179-39233,39238-39292,39297-39351,39356-39410,39415-39469,39474-39528,39533-39587,39592-39646,39651-39705,39710-39764,39769-39823,39828-39882,39887-39941,39946-39990,40000-40054,40059-40113,40118-40172,40177-40231,40236-40290,40295-40349,40354-40408,40413-40467,40472-40526,40531-40585,40590-40644,40649-40703,40708-40762,40767-40821,40826-40880,40885-40939,40944-40998,41003-41057,41062-41116,41121-41175,41180-41234,41239-41293,41298-41352,41357-41411,41416-41470,41475-41529,41534-41588,41593-41647,41652-41706,41711-41765,41770-41824,41829-41883,41888-41942,41947-42001,42006-42060,42065-42119,42124-42178,42183-42237,42242-42296,42301-42355,42360-42414,42419-42473,42478-42532,42537-42591,42596-42650,42655-42709,42714-42768,42773-42827,42832-42886,42891-42945,42950-42994,43000-43054,43059-43113,43118-43172,43177-43231,43236-43290,43295-43349,43354-43408,43413-43467,43472-43526,43531-43585,43590-43644,43649-43703,43708-43762,43767-43821,43826-43880,43885-43939,43944-43998,44003-44057,44062-44116,44121-44175,44180-44234,44239-44293,44298-44352,44357-44411,44416-44470,44475-44529,44534-44588,44593-44647,44652-44706,44711-44765,44770-44824,44829-44883,44888-44942,44947-44991,45000-45054,45059-45113,45118-45172,45177-45231,45236-45290,45295-45349,45354-45408,45413-45467,45472-45526,45531-45585,45590-45644,45649-45703,45708-45762,45767-45821,45826-45880,45885-45939,45944-45998,46003-46057,46062-46116,46118-46172,46177-46231,46236-46290,46295-46349,46354-46408,46413-46467,46472-46526,46531-46585,46590-46644,46649-46703,46708-46762,46767-46821,46826-46880,46885-46939,46944-46998,47003-47057,47062-47116,47118-47172,47177-47231,47236-47290,47295-47349,47354-47408,47413-47467,47472-47526,47531-47585,47590-47644,47649-47703,47708-47762,47767-47821,47826-47880,47885-47939,47944-47998,48003-48057,48062-48116,48118-48172,48177-48231,48236-48290,48295-48349,48354-48408,48413-48467,48472-48526,48531-48585,48590-48644,48649-48703,48708-48762,48767-48821,48826-48880,48885-48939,48944-48998,49003-49057,49062-49116,49118-49172,49177-49231,49236-49290,49295-49349,49354-49408,49413-49467,49472-49526,49531-49585,49590-49644,49649-49703,49708-49762,49767-49821,49826-49880,49885-49939,49944-49998,50003-50057,50062-50116,50118-50172,50177-50231,50236-50290,50295-50349,50354-50408,50413-50467,50472-50526,50531-50585,50590-50644,50649-50703,50708-50762,50767-50821,50826-50880,50885-50939,50944-50998,51003-51057,51062-51116,51118-51172,51177-51231,51236-51290,51295-51349,51354-51408,51413-51467,51472-51526,51531-51585,51590-51644,51649-51703,51708-51762,51767-51821,51826-51880,51885-51939,51944-51998,52003-52057,52062-52116,52118-52172,52177-52231,52236-52290,52295-52349,52354-52408,52413-52467,52472-52526,52531-52585,52590-52644,52649-52703,52708-52762,52767-52821,52826-52880,52885-52939,52944-52998,53003-53057,53062-53116,53118-53172,53177-53231,53236-53290,53295-53349,53354-53408,53413-53467,53472-53526,53531-53585,53590-53644,53649-53703,53708-53762,53767-53821,53826-53880,53885-53939,53944-53998,54003-54057,54062-54116,54118-54172,54177-54231,54236-54290,54295-54349,54354-54408,54413-54467,54472-54526,54531-54585,54590-54644,54649-54703,54708-54762,54767-54821,54826-54880,54885-54939,54944-54998,55003-55057,55062-55116,55118-55172,55177-55231,55236-55290,55295-55349,55354-55408,55413-55467,55472-55526,55531-55585,55590-55644,55649-55703,55708-55762,55767-55821,55826-55880,55885-55939,55944-55998,56003-56057,56062-56116,56118-56172,56177-56231,56236-56290,56295-56349,56354-56408,56413-56467,56472-56526,56531-56585,56590-56644,56649-56703,56708-56762,56767-56821,56826-56880,56885-56939,56944-56998,57003-57057,57062-57116,57118-57172,57177-57231,57236-57290,57295-57349,57354-57408,57413-57467,57472-57526,57531-57585,57590-57644,57649-57703,57708-57762,57767-57821,57826-57880,57885-57939,57944-57998,58003-58057,58062-58116,58118-58172,58177-58231,58236-58290,58295-58349,58354-58408,58413-58467,58472-58526,58531-58585,58590-58644,58649-58703,58708-58762,58767-58821,58826-58880,58885-58939,58944-58998,59003-59057,59062-59116,59118-59172,59177-59231,59236-59290,59295-59349,59354-59408,59413-59467,59472-59526,59531-59585,59590-59644,59649-59703,59708-59762,59767-59821,59826-59880,59885-59939,59944-599





```

QY      356   CTVAGMEEEFVCLKQCICPHIAYEAGESSYWRRIYLEGASAKYQCHSGSLPNGODTYCTE 415
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       135   CTAQGMEEFEVCVRKCFHYHVENGDSAWKEIKTIYYQGSLKYQCIINGSYLSQNGDLMTC 194

QY      416   NGMSPPPKCVRIRK 428
          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       195   NGMSPPPKCIRIK 207

RESULT    6

apollipoprotein H-related protein 23L1 - mouse
H35068
C:Species: Mus musculus (house mouse)
C:Date: 27-Jul-1990 #sequenceRevision 31-Dec-1993 #textChange 05-Nov-1999
C:Accession: H35068; A35069; B35069; I35068
R:VIK, D.P.; Munoz-Ganoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.
J. Biol. Chem. 265, 3193-3201, 1990
A:Title: Identification and sequence analysis of four complement factor H-related transmembrane proteins from mouse
A:Reference number: A35070; MUID:90153969
A:Accession: H35068
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-303 <VIK>
A:CROSS-references: GB:M29007; NID:g192557; PIDN:AAA37413.1; PID:g309164; GB:J05259
A>Note: Translation of the nucleotide sequence is not complete
C:Superfamily: Complement factor H repeat homology
C:Keywords: duplication
F:28-81/Domain: complement factor H repeat homology <FH01>
F:86-146/Domain: complement factor H repeat homology <FH02>
F:150-203/Domain: complement factor H repeat homology <FH03>
F:212-266/Domain: complement factor H repeat homology <FH04>
```

Query Match	33.0%	Score 811	DB 2	Length 303
Best Local Similarity	70.8%	Pred. No. 1.1e-49		
Best Match 136	Conservative 19	Mismatches 37	Indels 0	Gaps 0

[illegible]

RESULT 7  
 S46199  
 Probable complement regulatory plasma protein SBI - barred sand bass  
 C:Species: Parablax neblifer (barred sand bass)  
 C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 08-Oct-1999  
 C:Accession: S46199; 577894  
 R:Dahmen, A.; Kaidoh, T.; Zipfel, P.F.; Gagli, I.  
 Biochem. J. 301, 391-397, 1994  
 A:Title: Cloning and characterization of a cDNA representing a putative complement-regu  
 A:Reference number: S46199; MUID:94310039  
 A:Accession: S46199  
 A:Molecule type: mRNA  
 A:Residues: 1-1053 <DAH1>  
 A:Cross-References: EMBL:L21703; NID:G639894; PIDN:AAA92556.1; PID:G639895  
 A:Experimental source: liver  
 A:Accession: S77894  
 A:Molecule type: protein  
 A:Residues: 526-532, 'X', 534-537, 809-817, 'X', 819-826 <DAH2>  
 C:Genetics:

A:Gene: SB1  
C:Superfamily: complement factor H repeat homology  
C:Keywords: glycoprotein  
F:89-145/Domain: complement factor H repeat homology <FH01>  
F:334-389/Domain: complement factor H repeat homology <FH02>  
F:450-502/Domain: complement factor H repeat homology <FH03>  
F:568-654/Domain: complement factor H repeat homology <FH04>  
F:682-738/Domain: complement factor H repeat homology <FH05>  
F:743-802/Domain: complement factor H repeat homology <FH06>  
F:935-989/Domain: complement factor H repeat homology <FH07>  
F:993-1052/Domain: complement factor H repeat homology <FH08>

Query Match	23.2%	Score 569.5	DB 2	Length 1053
Best Local Similarity	26.7%	Pred No. 3.1e-32		
Matches 148; Conservative	60;	Mismatches 169;	Indels 177;	Gaps 19

QY	21	EQLSSEGOATYKCRHPGRTGTLVKKCKNGBWSPNSBSRJCRRKRPDGGHDPFGSFLR	80
Dd	47	EASVYGGQHVAVGNCNVGS--GFFKLVCVECKMETRGAK--CQPSRGHGDAQAFADFHL	1020
QY	81	AVGSFEFGAAVYTTCDGEGYLLCEIDYREDDDAGQTNNDIPICVAVKCLVYTELENGRIY	1400
Dd	103	AEGNDPEVGSKVYVYTCQGYOMVRINRYRCVAGMGVGVAVPVCESQOC-PLIHNNVQY	1610
QY	141	SGAAPDDEYFEGOVVRECNSSGFKI--EGQKEMHCSENGLMSNKPCCVEISCLPPVEN	1990
Dd	162	IGGPE--EATFGVNVRSCKSRSEILDGSELYCDEKDMSGFVYPCAKAITCAIPIEN	2180
QY	200	GDGILKPEVYKENERFOYKCKQGEVYKEREADVCTGSQ---NNOPSCSEMTCL-----	2500
Dd	219	GNVPBIAIREYKENVDLHNECDRAKRIHDR--PSTCIKQIAKEMSPPLPICSIKRILTIMD	2770
QY	251	-TPYLP-----NCITY-----	2600
Dd	278	GTRYEPARVNLFPSEBTLKVICARTSWISTPOETSVYTTCCDNCNEMSIIRPTCEVRCSNR	3370
QY	261	-PHRI-----KRIIDDEIRBECKNGEFPATRSPSKCTINGWLPAP-----	3000
Dd	338	REHVDMSDVNRSMERYTLLDDMTIRWCXKRGVYRTGGVYTWATGRNGMMNPLCEVYKTSKE	3970
QY	301	-----RCSLPCDCFPQFKH	3140
Dd	398	NIGDAVIGTDKQIYLNLMOKAIYACGBESNRRIITLTGSENWSGDRCYAKPCLPLPKDP	4570
QY	315	GLUYEESRRPYPVPIGKEYSYICDNGETTPRSQSYWYDLRCTYVNGEPEVYCLRQCI--	3720
Dd	458	NGFF--RGPY--TGRVLYTTCCKDYLKFLTEGMAWEACVGVW---PELTTCISN	5050
QY	373	-FHYVEGESSYMORRIIEGOSAAVOCHSGSYSLPNGODTYTTCENGW-----	4180
Dd	506	TTQGFPELIPNAEVI--RRPYEVQTVAVICNGYS--TOANSFSCENGMLLYGLSPDQ	5600
QY	419	-----SPPK	4230
Dd	561	ICTLADNCGPPE	5740

RESULT 8  
D35069  
complement factor H-related protein 9c4 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 27-Jul-1990 #sequence\_revision 31-Dec-1993 #text\_change 20-Aug-1999  
C:Accession: D35069; A35070; B35070; C35070; D35070; E35069; F35069; F35070;  
R:Vik, D.P.; Munoz-Canoves, P.; Kozono, H.; Martin, L.G.; Tack, B.F.; Chaplin, D.D.  
J. Biol. Chem. 265, 3193-3201, 1990  
A:Title: Identification and sequence analysis of four complement factor H-related tra  
A:Reference number: A35070; MUID:90153969  
A:Accession: D35069  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-808 <Vik>  
A:Cross-references: GB:M29009; NID:g192563; PIDN:AAA37416.1; PID:g309166; GB:J05259





A:Reference number: A43023; MUID:91217619  
A:Contents: annotation: exon-intron boundaries  
R:Chung, L.P.; Gagnon, J.; Reid, K.B.M.  
Mol. Immunol. 22, 427-433, 1985  
A>Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequence as-  
ced by cyanogen bromide treatment.  
A:Reference number: A93134; MUID:85296001  
A:Accession: A93134  
A:Molecule type: protein  
A:Residues: 49-81 <CH>  
A>Note: this paper reports amino-terminal sequences of the intact protein and of a bound  
R:Hessing, M.; Katers, D.; Takeya, H.; van't Veer, C.; Hackney, T.M.; Iwanaga, S.; Bont  
FBS Lett. 317, 228-232, 1993  
A>Title: The region Ser1333-Arg356 of the alpha-chain of human C4b-binding protein is  
A:Reference number: S29492; MUID:93146164  
A:Accession: S29492  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 381-404 <HS>  
R:Suzuki, K.; Nishioke, J.  
J. Biol. Chem. 263, 17034-17039, 1988  
A>Title: Binding site for vitamin K-dependent protein S on complement C4b-binding protein  
A:Reference number: A31785; MUID:89034204  
A:Accession: A31785  
A:Molecule type: protein  
A:Residues: 495-505, 'X', 507-510, 'X', 512-515 <SV>  
A>Note: this peptide appears to bind protein S  
R:Dahlback, B.; Smith, C.A.; Muller-Eberhard, H.J.  
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983  
A>Title: Visualization of human C4b-binding protein and its complexes with vitamin K-dep-  
A:Reference number: A93950; MUID:83221615  
A:Contents: annotation: electron microscopy: three-dimensional structure; ligand binding  
R:Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Nino, Y.  
Biochem. Biophys. Res. Commun. 174, 222-227, 1991  
A>Title: Genomic organization of the alpha chain of the human C4b-binding protein gene.  
A:Reference number: I52244; MUID:91113199  
A:Accession: I52244  
A>Status: translation not shown; translated from GB/EMBL/DDJ  
A:Molecule type: DNA  
A:Residues: 1-597 <AS>  
A:Cross-references: GB:M62486; NID:g190496; PIDN:AAA36506.1; PID:g190500  
C:Comment: C4BP controls the classical pathway of complement activation. It binds as a  
C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla  
C:Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment C4  
disulfide bonds. The beta chain binds the vitamin K-dependent plasma protein S. A minor f  
actors V and VIII.  
C:Comment: The molecule has a central body supporting seven tentacles (alpha chains), ea  
C:Genetics:  
A:Gene: GDB:C4BPA  
A:Cross-references: GDB:120568; OMIM:120830  
A:Map position: 1932-1932  
A:Introns: 48/1, 110/1, 143/2, 172/1, 236/1, 297/1, 362/1, 425/1, 482/1, 540/3  
C:Complex: octamer of seven alpha chains and one beta chain  
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla  
F:1-48/Domain: signal sequence #status predicted <SIG>  
F:49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>  
F:50-108/Domain: complement factor H repeat homology <FH1>  
F:113-170/Domain: complement factor H repeat homology <FH2>  
F:175-234/Domain: complement factor H repeat homology <FH3>  
F:239-294/Domain: complement factor H repeat homology <FH4>  
F:299-360/Domain: complement factor H repeat homology <FH5>  
F:364-422/Domain: complement factor H repeat homology <FH6>  
F:381-404/Region: complement C4b binding #status predicted  
F:426-480/Domain: complement factor H repeat homology <FH7>  
F:484-538/Domain: complement factor H repeat homology <FH8>  
F:521,506,528/Binding site: carbohydrate (Asn) (covalent) #status experimental

OY	2	DCKGPPRPNSSIIISGSWSQLYSEGTQATVYCGRGY-PTLTITVKVC-KNEWSPSNS	59
Db	49	NCGPRLTSLFAAPMIDILTEFRFKTGTTLLKYKCLDPGVASHSTQTULTCSNDSGBMYNT--	106
OY	60	RICKRRPCGHGPDTPFGSFRLLAVGSEFEFGAAVVYTCDGTYQLLEIDYRECDAD-----G	115
Db	107	-FCIARRKHRRHPELNRNG-QVEIKTKDLSFSGSIIEFSCSDGFPLIGSTSR-CEVDQRVG	162
OY	116	WNNDIPICEVNYCLPYTELENGRIYSGAEPDOEYYFGQVVRPECNSGRIEGOKEMHS	175
Db	163	WSHLPDCEIVAKCKRPDI RNCR-----HSGENFTAYGSVITYSCDPRSLIGHASICT	218
OY	176	-EN--GLMSNEKPQOCEVLICLDPRYENGDIY-LKPVYKENERFOYKCKOGFYVERKD	230
Db	219	VENETITGWRRSPPIKECIKTCKRPDYSHGEMVSGFCPIYNYKDTIFYEKCKGFLV--RGS	276
OY	231	AV--C-TGSGMWPO-PSCSEMTCLT-PTYINGIYT--PHRIK--HRIDEIRYECKNF	280
Db	277	SYIHCDADASKMPSPACPNSCINLPDI PHASMETYPRPTKEDVYVGTVLKRYCHPGY	336
OY	281	YVATRSVPVKCTITGG-WIPAPRCSLKPCDPFOFKRGRLYEESRRP-----	325
Db	337	KPTTEDPTTVICQKNLRMTPIYGCEALCCPERKLNNGETITHKKSRAHHCVFFYGDDEIS	396
OY	326	-----YPPVPIG-----KEYSYCDN	341
Db	397	FSCHEHSRFSALCGDGDTWSPRTPSGCDICNPFPKI AHGHYKQSSYSFPEKEELIYECOK	456
OY	342	GTTTSQSQTWDYLRCITVGNWBEPEVP-CLRQCIFHYIYEYESSTWMORYITEGSAKYQCHS	400
Db	457	GYILVQA---KLSCSYSHMSAPAPQCKALKCRKPDELNGRLSYDKDQYVEPENVTIQCD	513
OY	401	GYSLPNGODTFYCTEN-GMSPP-PKC	424
Db	514	GYGVV-GPOSTICSGNKRTWYPEYPKC	538
RESULT 11			
146001			
Cab-binding protein alpha chain - bovine			
C.Species: Bos primigenius taurus (cattle)			
C.Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text change 10-Sep-1999			
C.Accession: I46001; s43190			
J.Hilliarp, A.; Thern, A.; Dahlback, B.			
J.Immunol. 153, 4190-4199, 1994			
A.Title: Bovine Cab binding protein. Molecular cloning of the alpha- and beta-chains			
A.Reference number: I46001; MUID:95015909			
A.Accession: I46001			
A.Status: preliminary; translated from GB/EMBL/DBJ			
A.Molecule type: mRNA			
A.Residues: 1-610 <HTML>			
A.Cross-references: EMBL:Z31693; NID:g469117; PTDN:CAA83498.1; PID:g469118			
C.Superfamily: Cab-binding protein alpha chain; complement factor H repeat homology			
F:50-107/Domain: complement factor H repeat homology <FH1>			
F:112-169/Domain: complement factor H repeat homology <FH2>			
F:174-234/Domain: complement factor H repeat homology <FH3>			
F:239-294/Domain: complement factor H repeat homology <FH4>			
F:299-362/Domain: complement factor H repeat homology <FH5>			
F:366-425/Domain: complement factor H repeat homology <FH6>			
F:429-483/Domain: complement factor H repeat homology <FH7>			
F:487-541/Domain: complement factor H repeat homology <FH8>			
Query Match 19.3%; Score 474.5; DB 1; Length 610;			
Best local Similarity 27.2%; Pred. No. 7, 2e-26;			
Matches 140; Conservative 76; Mismatches 180; Indels 119; Gaps 29;			
OY	5	GPPRENSELISGSWSQLYSEGTOATYKCRPGRTILGTYACKN-----GEWPPSN	57
Db	51	GIPPYLDFAPINELNETREFGTTLTYTCRPGYR-----ISSRNKLIDCGTDNW--K	102
OY	58	PSKICRRPCGHGDPFGSFRLAVGSEFEFGAAVVYTCDGTYQLLEIDYRECDADG-	115

```

Db 103 YKEFCYKXKCEBNPELLNG--QVIYKTDVXSFGEISEFSCSEGVLLIGSANSYQLODKCV 160
QY 116 -ATNPIDPIEYVKCLPYLELLENGRIYSGAEPDQETV-FGVVRFECNSGFKITEGCKEMH 173
Db 161 WMSDLPLOCTIKACEPPITISNGRHHNG--DEDEYTGSSIVYSCDDDFSMIGKASIS 216
QY 174 CS-EN--GLAMSEKPOCUEISCLPVRVENG--DBEYLKPYKENEROYCKOKGQFVK 226
Db 217 CKEVKEKITGWSMSPSPCKKAVICVQPVYVDGKITSG--FGPIYTYQOSTIVYACNKGFE-R 272
QY 227 ERGDV--C-TGSGMN-POPSCEMTCL-TPYIPNGIYTPHRIKIRIDE-----I 272
Db 273 LEGDSLHICEANSMNWPPEPCETELNGCLGLPHIPLHM--ERYDHQTOQEQQVYIDIGFLV 330
QY 273 RECKANGFPAIRSRYSKCTIT-----GWIPARCSLAKCPDPQF-KHG-RUYEESRR 324
Db 331 STCKHFGKRPEDGPT--TYTCQSNLEWSPYTECKEYCCPPEPNNLVYGSITLHRRPSTS 387
QY 325 PYFPPIGKEXSYVYCDNG-----F 343
Db 368 THCYTISDGIKISYECHSKYMPDALCTKHGTHMSPRTRPECKRPDCKSPRYLAHGCHKVYSKFE 447
QY 344 TTPSOSYWD-----YLRCTVNGWNEPEVP-CLROCIPIHYVEYGGSSYWMORRYIEG 391
Db 448 TFDHQAAYVECDMGYILVGAKEIJSSTSSGMSPAVPOCKALCLKPELEYGLSVLEKRYIYEP 507
QY 392 QSAKVQCHSGSYSLPNCQDPTVYCTEN-GMSPP-PKC 424
Db 508 ELITIQCESGYSIV-GSENIJTCSEDRTPWPEYVPC 541

```

RESULT 12

A45222

complement factor H-related protein DOWN16 precursor - human

C:Species: Homo sapiens (man)

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Mar-2000

C:Accession: A45222; S29609

R:Skeřxa, C.; Kuhh, S.; Gunther, K.; Lingelbach, K.; Zipfel, P.F.

J. Biol. Chem. 268, 2904-2908, 1993

A:Title: A novel short consensus repeat-containing molecule is related to human complement factor H

A:Reference number: A45222; MUID:93155112

A:Accession: A45222

A:Molecule type: mRNA

A:Residues: 1-331 <SKE>

A:Cross-references: EMBL:X68679; NID:g30869; PID:g30870

A:Experimental source: liver

A>Note: sequence extracted from NCBI backbone (NCBIP:124269)

C:Genetics:

A:Gene: GDB:HFL4

A:Cross-references: GDB:137201

C:Superfamily: complement C3d/Epstein-Barr virus receptor; complement factor H repeat h

C:Keywords: glycoprotein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-331/Product: complement factor H-related protein DOWN16 #status predicted <MAT>

F:23-83/Domain: complement factor H repeat homology <FH01>

F:87-141/Domain: complement factor H repeat homology <FH02>

F:147-204/Domain: complement factor H repeat homology <FH03>

F:211-265/Domain: complement factor H repeat homology <FH04>

F:269-330/Domain: complement factor H repeat homology <FH05>

F:108,186,206,310/Binding site: carbohydrate (asn) (covalent) #status predicted

[illegible]

QY	412	YCTENGWSPPPKCVR	426
		: :	
Db	129	TCTENGWSPTPRCIR	143

RESULT	13
A45900	

complement C3b receptor type 2 long form precursor - mouse (fragment)  
C3SpecIES: Mus musculus (house mouse)

```
C:\Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 08-Oct-1999
C:\Accession: 145000 140300
```

R; Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.

A; Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2

A; Accession: A45900

A; molecule type: mRNA

A;Cross-references: GB:M36470

R;Kim, Y.U.; Kinoshita, T.; Molin

J. Exp. Med. 181, 131-139, 1995

A;Accession: 148306

```
A;status: preliminary
A;molecule type: DN
```

A;Residues: 21-36 / <RES>  
A;Cross-references: EMBL:U17128: NID:a595980: PTDN:

C;Genetics:  
A;Gene: Crr2

C/superfamily: complement factor H repeat  
A/introns: 80/1; 113/2; 142/1; 275/1; 333/1

F:83-140/Domain: complement factor H repeat homolog  
F:22-78/Domain: complement factor H repeat homolog

F;145-211/Domain:	complement factor H repeat	homology <FH0
E;217-272/Domain:	complement factor H repeat	homology <FH0

F;276-331/Domain:	complement factor H repeat	homology <FH0
E;336-394/Domain:	complement factor H repeat	homology <FH0

F;399-458/Domain: complement factor H repeat homology <FH0  
F:467-523/Domain: complement factor H repeat homology <FH0

F;531-587/Domain: complement factor H repeat homology <FH0

Query Match	Score	DB 2	Length
18.5%	453	676	

Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

QY 24 YSEGTQATYKCRPGYRTLGTIVKVCKNGEWVPSNP SRICRKRPCGHPGDTPEGSFRLAVG 83

Db 42 FAIGTWYKCRPGYFRKSFITCLETSKW--SDAQQFCRKRCMPQEPPLHGSVHINTG 99

0Y 84 SEFEFGAKVYTCDEGYQLLGEIDYRECDADG--WINDIPICEVVKCLPTELENGRIV 1400

Db 100 --IEFGSTITYSCNÖGYRLIGDSSATCIIVSDNTVMWDNDMPLECSIPCESPPAISNGDFY 157

0Y 141 SGAEPDOEYFQGVVRECSG-----FKIEGOKEMHCSEN-----GLWSNEKPOCV-E 189

Db 158 SSSRD---SEFYGMVTTYCHTGKNRKLEFDLVEKSIYCTSKDNOVGIIWNSPPCOTPR 214

190 ISCLPPRVENG---DGIYLKPVYKENERFOYKCKOGFVYKFRGDVVC-TGSGWNPQ-PSC 344

Db 215 VKCPMEPIENGIVESG--FKHSFEINDVTYFKCKSGETMKCSRIAWCCDNPKWSPDIPCC 273

OV 245 FEMTCTPY-TPNGTYTPHRTKHRTDDETPYFCKNGFVBPATBSPDYCKCTTTC-ET-BDD 301

Db 273 -EMGCI PPNITL HGDYNNKRDEEFSSVGCOKVSTGNCB -YET ICTNT VECSTI CMTGCMVPT 330

302 CSI.KBCD-FB-0FKUCBI VVEEBBNDNDTCKNVCVVDDVCTE - ENDOGENEUT PG 1CC

[illegible]

Db 331 CEVSCDAIPNHLHGHVFLPPN-----IQLGAEVSFVCDLGFQLKGPSS-----QC 378  
 QY 357 TVNG----WPEVPCLOCLFH---YVEYGESSYWRRIYEGQAKYQCHSGYSLPMDQ 409  
 Db 379 IPEGETYIMNNKRPVCSQISCDPPEVKNARKPYSLPIPVGLVRLRYTCSPSYRL-IGEK 437  
 QY 410 TYVC-TEN---GW-SPPKC 424  
 Db 438 AIFCISENYHATWDRKAPIC 458

RESULT 14

C4BP protein alpha chain precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 20-Aug-1999  
 C:Accession: S57953  
 R:Hallarp, A.; Thern, A.; Dahlback, B.  
 Submitted to the EMBL Data Library, July 1995  
 A:Description: Molecular cloning of rat C4b-binding protein alpha- and beta-chains: struc  
 A:Reference number: S57953  
 A:Accession: S57953  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-558 <HL>  
 A:Cross-references: EMBL:Z50051; NID:g899379; PID:CA90391.1; PID:g899380

C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 F:15-72/Domain: complement factor H repeat homology <FH1>  
 F:77-134/Domain: complement factor H repeat homology <FH2>  
 F:139-199/Domain: complement factor H repeat homology <FH3>  
 F:204-258/Domain: complement factor H repeat homology <FH4>  
 F:263-324/Domain: complement factor H repeat homology <FH5>  
 F:328-386/Domain: complement factor H repeat homology <FH6>  
 F:390-443/Domain: complement factor H repeat homology <FH7>  
 F:447-501/Domain: complement factor H repeat homology <FH8>

Query Match 18.3%; Score 449; DB 2; Length 558;  
 Best Local Similarity 26.7%; Pred. No. 4e-24;

Matches 137; Conservative 71; Mismatches 198; Indels 108; Gaps 24;

QY 5 GPPRESEILSGWSGSEGLYSEGTQATYKCRPGY-RTLGLTVKCKN-GEWVSPNSRIC 62  
 ||||| : : : : : ||||| : : : : :  
 Db 16 GPPEDLYALPASEMNTDESHLTLRNCRPGYSRASSSGLCKPLGKM---QINLAC 72  
 QY 63 RKRCGHGPTPGSGFRLAVGSEFGAKVYVTCDESYQLG-EDYRECDADG-WTND 119  
 || : || : : : : || : : : : : || : : : : :  
 Db 73 VKKSCRNPGLDNG--KVEVKTDFLFGSJLEFSGSEGYILIGSSYSTYCELDGKGVMSDP 130  
 QY 120 IPICEVVKCLPVTELENGRIVSGAEPDQEVYFGQVVRFCNSGFKIEGQEMHCS---- 175  
 :  
 Db 131 LPECVIAKCMPPDISNGK--HNGREEFFTYRSSVTYKCDPFTLLGNASITCTVYVK 187  
 QY 176 ENGLSMNEKQCVESICLPRVENDG-GIYLKPYKKENERFOYCKOGFYVKEGDAVCT 234  
 :  
 Db 188 TVGWMSSPPTCEIILCPWKKVLHGTINSFGKTKYKDSRFYCKQKGFVLRGSGVLIHCE 247  
 QY 235 GSG-WNPQSPCEEMTCL-TPYIPNG--IYTPHRIK---HRIDELRYECKNGFYPARTRSP 287  
 :  
 Db 248 ADGWSMSPVCELSNCDIDIPNALITSPRKEDYVPVGLVRLRYCIRGYPATRQRP 307  
 QY 288 VSKCTI-----TGWIPARCSLAKCPDFPQFHKGLIYEESSRP-----YFVPIGKEYSY 337  
 :  
 Db 308 M--TVICQKDLGWSMLRGCKEICCPVDPKSVRIQHEKAKHPNDCTYF--FGDEVSY 361  
 :  
 QY 338 YCDN-----GFTPSQSYWYDLR----- 355  
 :  
 Db 362 TCQNDIMLMTACKSDGWHNRTPSCHOSCEPPALAHGRYTKSSSY--YVNTQYTYEECE 419  
 QY 356 -----CTVNGKEPEVP-CLROCIEHYVEYEGSSYWRRIYEGQAKYQCHSGYS 403  
 :  
 Db 420 GYRLVGEATITSCWYSQWTPAAPQCKALCRKREIENGVLSTKQDYVETENVTICDSEGFV 479

QY 404 LPNGDITYCTENG-WSP-----PPKCVRI 427  
 :  
 Db 480 M-LDSQITSCSENGCTWIPKVSRCDEQVPKDCEHV 512

RESULT 15

complement receptor 1 - chimpanzee (fragment)  
 C:Species: Pan troglodytes (chimpanzee)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 31-Jan-2000  
 C:Accession: I36936; I36937  
 R:Birmingham, D.J.; Shen, X.P.; Hourcade, D.; Nickells, M.W.; Atkinson, J.P.  
 J. Immunol. 153, 691-700, 1994  
 A:Title: Primary sequence of an alternatively spliced form of CRI. Candidate for the  
 A:Reference number: I36935; MUID:94292799  
 A:Accession: I36936  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2014 <BIR1>  
 A:Cross-references: GB:I24920; NID:g551564; PID:g557725

A:Accession: I36937  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-397,1751-2014 <BIR2>  
 A:Cross-references: GB:I24921; NID:g557726; PID:g557727  
 C:Genetics:  
 A:Gene: CRI  
 C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 C:Keywords: duplication; glycoprotein; tandem repeat  
 F:18-74/Domain: complement factor H repeat homology <FH01>  
 F:79-136/Domain: complement factor H repeat homology <FH02>  
 F:141-207/Domain: complement factor H repeat homology <FH03>  
 F:213-268/Domain: complement factor H repeat homology <FH04>  
 F:272-328/Domain: complement factor H repeat homology <FH05>  
 F:333-391/Domain: complement factor H repeat homology <FH06>  
 F:396-462/Domain: complement factor H repeat homology <FH07>  
 F:470-526/Domain: complement factor H repeat homology <FH08>  
 F:529-586/Domain: complement factor H repeat homology <FH09>  
 F:722-778/Domain: complement factor H repeat homology <FH10>  
 F:1041-1107/Domain: complement factor H repeat homology <FH11>  
 F:1172-1228/Domain: complement factor H repeat homology <FH12>  
 F:1233-1291/Domain: complement factor H repeat homology <FH13>  
 F:1296-1362/Domain: complement factor H repeat homology <FH14>  
 F:1432-1489/Domain: complement factor H repeat homology <FH15>  
 F:1625-1681/Domain: complement factor H repeat homology <FH16>  
 F:1749-1815/Domain: complement factor H repeat homology <FH17>  
 F:1823-1879/Domain: complement factor H repeat homology <FH18>  
 F:1884-1940/Domain: complement factor H repeat homology <FH19>

Query Match 18.2%; Score 447; DB 2; Length 2014;  
 Best Local Similarity 30.2%; Pred. No. 2.2e-23;  
 Matches 137; Conservative 64; Mismatches 165; Indels 88; Gaps 32;

QY 27 GTOATYKCRPEYRILGTVKCKNCEWVSPNSRICRRKRPCHGHPDTPFGSFRILAVGSEF 86  
 ||||| : : : : : ||||| : : : : :  
 Db 941 GTSUKYECRPYEGRPESITGLDNLW--SSPKDVCRRKSKCTPPDPVNG--MVAHVIDI 996  
 QY 87 EFGAKVYVTCDEGYQLDEIDYRECDADG---WTNDIPICEVYKCLPVTELENGRIVS 141  
 :  
 Db 997 QVGSRLNSCTTGHRLLH--SSAECILSGNSAHMSTRPICOIRIPCGLPPI-IANGDIS 1054  
 QY 142 GAABPDQEVYFGQVVRFCNSG-----FKIEGQEMHCSN---GLMSNEKPOC-VEI 190  
 :  
 Db 1055 TNRE---NFHYGSVYVTCNLSGSRKRYFELVGEPSIYCTSNDDYGVIMSGPAPQCIIPN 1111  
 QY 191 SCLPRVENG---DGILKPYKKENERFOYCKOGFYVKEGDAVCTG-SGWNPQ-PSG 244  
 :  
 Db 1112 KCTPPNENGLIVSDN--KSLFSLNEVVERFCQGFPMKGRHRYKCALAKMKEPELPSG 1168  
 QY 245 EEMTCLPYIPNGIYTP-HRIKTHRIDELRYECKNGFYPARTRSPVSKCTING-WIP-APR 301  
 :  
 Db 1169 SRVCOPPPEILHGEHTPHQDNFSPGQEVFYVSGEBG-YDLGAGASLHCTPGGDMSPAPR 1227

Tue Nov 21 16:57:31 2000

us-09-316-163-14.rpr

Page 10

QY	302	CSLPCPC--PPOFHGHLYEESRRPFP--VIGKEYSYCDNFTFPPSGSYMDYLHCT	357
Db	1228	CAVACSCDDPLQGLPHGRV-----LFLPLNQLGAKKVSFYVDEBDFRLKGS--VSHCV	1276
QY	358	VNG-----WEDEPCLRLRCIF-----HX-VBYGESSYQWRRTIEGQSAKVCQ--H	399
Db	1277	LVGKRLSLNNNSVPCBQ--ILCPPLPALINGRHGTGTFQDILP-----GKEISTYCDHP	1328
QY	400	SGISLP---NGQDTTYCTE---NG-W-SPPKRC	424
Db	1329	PDRCMTNLIGESTIRCTSDPHSGNWSSPAPRC	1362

Search completed: November 21, 2000, 16:47:45  
Job time: 374 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:49:24 ; Search time 27.94 seconds

(without alignments)  
489.396 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454

Sequence: 1 EDCKGPPRENSSEILSGSMS.....PTYCTENGWSPPCVRIRK 428

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2055	83.7	1234	1	CFRH_MOUSE
2	1714	69.8	1231	1	CFRH_MOUSE
3	481.5	19.6	597	1	C4BP_HUMAN
4	474.5	19.3	610	1	C4BP_BOVIN
5	466.5	19.0	331	1	CFHD_HUMAN
6	449	18.3	558	1	C4BP_RAT
7	437	17.8	2039	1	CRL_HUMAN
8	384.5	15.7	469	1	C4BP_MOUSE
9	373	15.2	1025	1	CR2_MOUSE
10	368	15.0	1033	1	CR2_HUMAN
11	336	13.7	661	1	F13B_HUMAN
12	332	13.5	668	1	F13B_MOUSE
13	335	12.8	330	1	CFHL_HUMAN
14	314.5	12.8	263	1	VCP_VACCV
15	310.5	12.7	768	1	LEM3_MOUSE
16	310	12.6	270	1	CFHE_HUMAN
17	309.5	12.6	830	1	LEM3_HUMAN
18	298.5	12.2	377	1	MCP_HUMAN
19	296	12.1	768	1	LEM3_RAT
20	290	11.8	610	1	LEM2_HUMAN
21	282.5	11.5	769	1	LEM3_SHEEP
22	277.5	11.3	611	1	LEM2_CANFA
23	272	11.1	612	1	LEM2_MOUSE
24	271.5	11.1	345	1	APDH_MOUSE
25	267.5	10.9	345	1	APDH_HUMAN
26	267.5	10.7	507	1	DAF_CAVPO
27	262.5	10.5	345	1	APDH_CANFA
28	256.5	10.3	551	1	LEM2_RABIT
29	255	10.4	360	1	CCPH_HSVSA
30	252	10.3	484	1	LEM2_PIG
31	246.5	10.0	646	1	LEM3_BOVIN
32	244.5	10.0	381	1	DAF_HUMAN
33	243	9.9	390	1	DAF1_MOUSE

34	241.5	9.8	958	1	HIG_DROME
35	238.5	9.7	340	1	DAF_PONPY
36	238.5	9.7	345	1	APDH_BOVIN
37	238	9.7	549	1	LEM2_RAT
38	236.5	9.6	407	1	DAF2_MOUSE
39	221.5	9.0	485	1	LEM2_BOVIN
40	218	8.9	297	1	APDH_RAT
41	195	7.9	372	1	LEM1_MOUSE
42	191.5	7.8	1019	1	LEF_TACTR
43	190.5	7.8	1019	1	LEF_CARRO
44	189.5	7.7	317	1	VB05_VACCC
45	189	7.7	317	1	VB05_VACCO

## ALIGNMENTS

```

RESULT 1
CFRH_MOUSE
ID      CFRAH_MOUSE      STANDARD:      PRT: 1234 AA.
AC      P06909;
DT      01-JAN-1988 (Rel. 06, Created)
DT      01-JAN-1988 (Rel. 06, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      COMPLEMENT FACTOR H PRECURSOR (PROTEIN BETA-1-H).
CN      CRH.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE: 8623353.
RA      Kristensen T., Tack B.F.;
RT      "Murine protein H is comprised of 20 repeating units, 61 amino acids
RT      in length."
RL      Proc. Natl. Acad. Sci. U.S.A. 83:3963-3967(1986).
RN      [2]
RP      SEQUENCE OF 1-19 FROM N.A.
RX      STRAIN-BALB/C.
RX      MEDLINE: 90148935.
RA      Munoz-Canoves P., Tack B.F., Vik D.P.;
RT      "Analysis of complement factor H mRNA expression: dexamethasone and
RT      IFN-gamma increase the level of H in L cells."
RL      Biochemistry 28:9891-9897(1989).
RN      [3]
RP      SEQUENCE OF 1-18 FROM N.A.
RX      MEDLINE: 90111033.
RA      Natsume-Sakai S., Nonaka M., Harada Y.N., Shreffler D.C.,
RA      Moriaki K.;
RT      "Demonstration of an unusual allelic variation of mouse factor H by
RT      the complete cDNA sequence of the H.2 allotype."
RL      J. Immunol. 144:358-362(1990).
CC      -1- FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF
CC      C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE
CC      C3BB COMPLEX (C3 CONVERTASE) AND THE (C3b)NBB COMPLEX (C5
CC      CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.
CC      -1- POLYMORPHISM: TWO CODOMINANT ALLELES OF FACTOR H ARE PRESENT IN
CC      MICE.
CC      -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL: M12660; AAA37759.1; -
DR      EMBL: J02891; AAA37795.1; -
DR      EMBL: M31979; AAA37762.1; -
DR      PIR: A26154; NEMSH.
DR      HSP: P08603; IHFI.

```

FT	CABOBYD	1061	1061	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CABOBYD	1225	1225	N-LINKED (GLCNAC. . .) (POTENTIAL)
SO	SEQUENCE	1234 AA:	139082 MW:	CSAC02F341B957E7 CRC64:
	Query Match	83.7%;	Score 2055;	DB 1; Length 1234;
	Best Local Similarity	82.9%;	Pred. No. 7, 2e-149;	
	Matches 355;	Conservative 28;	Mismatches 45;	Indels 0; Gaps
QY	1	EDCGPPRENSSELLSGWSWOLYSEGTOKMYKCRPEYRFLGTVIVCKNGKGVWPSNPSR	60	
DB	19	EDCGPPPRENSSELLSGWSWOLYSEGTOKMYKCRPEYRFLGTVIVCKNGKGVWASNPSSR	78	
QY	61	ICRRPCGHPEDTFFGSGFRLAVGSEFEFGAKVYVTCDEGYOLLGEIDYRECDADGWTNDI	120	
DB	79	ICRRKPCGHPEDTFFGSGFRLAVGSGDFEFGAKVYVTCDDGYOLLGEIDYRECDADGWTNDI	138	
QY	121	PICEVYKCLPTELENGRIVSGAAEPDQEVYFGQVYAFECNSGFKIEGKCEHNCSENGIM	180	
DB	139	PLCEVYKCLPTELENGRIVSGAAETDQEVYFGQVYAFECNSGFKIEGKCEHNCSENGIM	198	
QY	181	SNKRPOCEVIECLPPRVENGDIILKLVYKENERFQYKQKGQFYKKEKRDVATGSGWNP	240	
DB	199	SNKRRCVEILCTPPRVENGDIIVKRVYKENERHYHKCKHGVYKPKRDVATGSGWMS	258	
QY	241	OPSCHEMTCLTPYIPNGLIYTPHRIKXHNIDEIRIECKNGNGYIPATRSVSKCTITGMIPAR	300	
DB	259	OPFCEKRSKSPYLLINGIYTPHRIIHSDEIRIECYNGYIPVYGVSVSKCTPTGMIPVP	318	
QY	301	RCSLKPDPFOPFKGRILYKESRRPYPVPVIGKESYSYCNNGFPTPSQSWDYLRCITVNG	360	
DB	319	RCTLKPCEFPQPKGRILYKESLRLNPNFVSIQNKYSYKCONGSEPPSGYSMDYLRCITVNG	378	
QY	361	WEPEVPCLRQCIPIHYVEGESSYVQRRYIESQASAKVOCBSGSLPNCQDTTYCTLENGMSP	420	
DB	379	WEPEVPCVRKCVFHYVENGDSAVMEKVVQSGSLKVOCYNGYSLQNGQDTMTCTENGWSP	438	
QY	421	PKCVYRK 428		
DB	439	PKCIRIK 446		
RESULT 2				
CPAH_HUMAN				
ID	CPAH_HUMAN	STANDARD;	PRT;	1231 AA.
AC	01-AUG-1988	(Rel. 08, Created)		
DT	01-JAN-1990	(Rel. 13, Last sequence update)		
DT	30-MAY-2000	(Rel. 39, Last annotation update)		
DE	COMPLEMENT FACTOR H PRECURSOR.			
GN	HF1 OR HF OR CFH.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RC	TISSUE=LIVER;			
RX	MEDLINE: 88134059.			
RA	Ripoche J., Day A.J., Harris T.J.R., Sim R.B.;			
RT	"The complete amino acid sequence of human complement factor H.;"			
RL	Biochem. J. 249:593-602(1988).			
RN	[2]			
RN	SEQUENCE OF 53-445 FROM N.A.			
RA	MEDLINE: 87054207.			
RX	Schulz T.F., Schwable W., Stanley K.K., Weiss E., Dierich M.P.;			
RT	"Human complement factor H: isolation of cDNA clones and partial cDNA			
RT	sequence of the 38-kDa tryptic fragment containing the binding site			
RT	for C3b.;"			
RL	Eur. J. Immunol. 16:1351-1355(1986).			
RN	[3]			
RN	SEQUENCE OF 226-445 FROM N.A., AND PARTIAL SEQUENCE.			
RA	MEDLINE: 86169701.			
RA	Kristensen T., Wetzel R.A., Tack B.F.;			

RT "Structural analysis of human complement protein H: homology with C4b  
RT binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RL J. Immunol. 136:3407-3411(1986).  
[4]  
RP SEQUENCE OF 1047-1231 FROM N.A.  
RX MEDLINE: 91201892.  
RA Estaller C., Koistinen V., Schwaebler W., Dierich M.P., Weiss E.H.;  
RT "Cloning of the 1.4-kb mRNA species of human complement factor H";  
RT reveals a novel member of the short consensus repeat family related  
RL to the carboxy terminal of the classical 150-kda molecule.;  
J. Immunol. 146:3190-3196(1991).  
[5]  
RP SEQUENCE OF 19-35.  
RX MEDLINE: 83048213.  
RA Sim R.B., Disciplo R.G.;  
RT "Purification and structural studies on the complement-system control  
RL protein beta 1H (factor H).";  
RL Biochem. J. 205:285-293(1982).  
[6]  
RP STRUCTURE BY NMR OF 927-985 (SUSHI 16).  
RX MEDLINE: 91278097.  
RA Norman D.G., Barlow P.N., Baron M., Day A.J., Sim B., Campbell I.D.;  
RT "Three-dimensional structure of a complement control protein module  
RL in solution.";  
J. Mol. Biol. 219:717-725(1991).  
[7]  
RP STRUCTURE BY NMR OF 264-322 (SUSHI 5).  
RX MEDLINE: 92232649.  
RA Barlow P.N., Norman D.G., Steinkasserer A., Horne T.J., Pearce J.,  
RL Driscoll P.C., Sim B., Campbell I.D.;  
RT "Solution structure of the fifth repeat of factor H: a second example  
RL of the complement control protein module.";  
J. Biochemistry 31:3626-3634(1992).  
[8]  
RP STRUCTURE BY NMR OF 866-985 (SUSHI 15 AND 16).  
RX MEDLINE: 93323119.  
RA Barlow P.N., Steinkasserer A., Norman D.G., Kieffer B., Wiles A.P.,  
RL Sim B., Campbell I.D.;  
RT "Solution structure of a pair of complement modules by nuclear  
RT magnetic resonance.";  
J. Mol. Biol. 232:268-284(1993).  
[9]  
RP FUNCTION: FACTOR H FUNCTIONS AS A COFACTOR IN THE INACTIVATION OF  
CC C3B BY FACTOR I AND ALSO INCREASES THE RATE OF DISSOCIATION OF THE  
CC C3BB COMPLEX (C3 CONVERTASE) AND THE (C3B)NMB COMPLEX (C5  
CC CONVERTASE) IN THE ALTERNATIVE COMPLEMENT PATHWAY.  
CC -1- SIMILARITY: CONTAINS 20 SUSHI (SCR) REPEATS.  
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION  
CC 341 ONWARD DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
DR EMBL: Y00716; CAA68704.1; -  
DR EMBL: X04697; CAB41739.1; ALT\_FRAME.  
DR EMBL: M65294; AAA35948.1; -  
DR PIR: S00254; NBH0H.  
DR PIR: S00254; S00254.  
DR PIR: S03013; S03013.  
DR PDB: IHCC; 1S-APR-92.  
DR PDB: IHH; 1S-JUL-93.  
DR PDB: IHH; 1S-JUL-93.  
DR MIM: 134370; -  
DR INTERPRO: IPR000436; -  
DR PFAM: PF00084; sushi. 20.  
KW Complement alternate pathway; Plasma; Glycoprotein; Repeat; Sushi;  
KW Signal; 3D-structure; Polymorphism.  
FT SIGNAL 1 18  
FT CHAIN 19 1231 COMPLEMENT FACTOR H.

FT	DOMAIN	20	1229	20 X SUSHI (SCR) REPEATS.
FT	REPEAT	20	81	SUSHI 1.
FT	REPEAT	84	142	SUSHI 2.
FT	REPEAT	145	206	SUSHI 3.
FT	REPEAT	209	263	SUSHI 4.
FT	REPEAT	266	321	SUSHI 5.
FT	REPEAT	324	386	SUSHI 6.
FT	REPEAT	388	443	SUSHI 7.
FT	REPEAT	447	506	SUSHI 8.
FT	REPEAT	508	565	SUSHI 9.
FT	REPEAT	568	624	SUSHI 10.
FT	REPEAT	629	685	SUSHI 11.
FT	REPEAT	690	745	SUSHI 12.
FT	REPEAT	752	804	SUSHI 13.
FT	REPEAT	810	865	SUSHI 14.
FT	REPEAT	869	927	SUSHI 15.
FT	REPEAT	930	985	SUSHI 16.
FT	REPEAT	988	1044	SUSHI 17.
FT	REPEAT	1047	1103	SUSHI 18.
FT	REPEAT	1108	1164	SUSHI 19.
FT	REPEAT	1166	1229	SUSHI 20.
FT	DISULFID	21	66	BY SIMILARITY.
FT	DISULFID	52	80	BY SIMILARITY.
FT	DISULFID	85	129	BY SIMILARITY.
FT	DISULFID	114	141	BY SIMILARITY.
FT	DISULFID	146	192	BY SIMILARITY.
FT	DISULFID	178	205	BY SIMILARITY.
FT	DISULFID	210	251	BY SIMILARITY.
FT	DISULFID	237	262	BY SIMILARITY.
FT	DISULFID	267	309	BY SIMILARITY.
FT	DISULFID	294	320	BY SIMILARITY.
FT	DISULFID	325	374	BY SIMILARITY.
FT	DISULFID	357	385	BY SIMILARITY.
FT	DISULFID	389	431	BY SIMILARITY.
FT	DISULFID	416	442	BY SIMILARITY.
FT	DISULFID	448	494	BY SIMILARITY.
FT	DISULFID	477	505	BY SIMILARITY.
FT	DISULFID	509	553	BY SIMILARITY.
FT	DISULFID	536	564	BY SIMILARITY.
FT	DISULFID	569	611	BY SIMILARITY.
FT	DISULFID	597	623	BY SIMILARITY.
FT	DISULFID	630	673	BY SIMILARITY.
FT	DISULFID	659	684	BY SIMILARITY.
FT	DISULFID	691	733	BY SIMILARITY.
FT	DISULFID	719	744	BY SIMILARITY.
FT	DISULFID	753	792	BY SIMILARITY.
FT	DISULFID	781	803	BY SIMILARITY.
FT	DISULFID	811	853	BY SIMILARITY.
FT	DISULFID	839	864	BY SIMILARITY.
FT	DISULFID	870	915	BY SIMILARITY.
FT	DISULFID	901	926	BY SIMILARITY.
FT	DISULFID	931	973	BY SIMILARITY.
FT	DISULFID	959	984	BY SIMILARITY.
FT	DISULFID	989	1032	BY SIMILARITY.
FT	DISULFID	1018	1043	BY SIMILARITY.
FT	DISULFID	1048	1091	BY SIMILARITY.
FT	DISULFID	1077	1102	BY SIMILARITY.
FT	DISULFID	1109	1152	BY SIMILARITY.
FT	DISULFID	1138	1163	BY SIMILARITY.
FT	DISULFID	1167	1218	BY SIMILARITY.
FT	DISULFID	1201	1228	BY SIMILARITY.
FT	CARBOHYD	529	529	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	718	718	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	822	822	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	882	882	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1029	1029	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1095	1095	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	402	402	H -> Y.
FT	CONFLICT	21	21	/FTID=VAR.001979.
FT	CONFLICT	30	30	C -> Q (IN REF. 3).
FT	CONFLICT			T -> V (IN REF. 3).

```

FT CONFLICT 34 34 T -> Q (IN REF. 3).
FT CONFLICT 53 54 RP -> IL (IN REF. 2).
FT STRAND 870 871
FT STRAND 876 876
FT STRAND 877 878
FT STRAND 879 881
FT STRAND 890 891
FT TURN 893 894
FT STRAND 896 900
FT STRAND 903 904
FT STRAND 907 907
FT STRAND 912 916
FT TURN 917 918
FT STRAND 919 920
FT STRAND 926 926
SQ SEQUENCE 1231 AA; 139125 MW; C65EC8CF800B3FD CRC64;

Query Match 69.8%; Score 1714; DB 1; Length 1231;
Best Local Similarity 67.5%; Pred. No. 6,4e-123;
Matches 289; Conservative 47; Mismatches 92; Indels 0; Gaps 0;

QY 1 EDCGKPPENSEILSGNSEQLTSGTQATYKCRPGYRTLTGTVKCKNGEWPNPSR 60
DB 19 EDCNELPKRNPTILLGMSNDQTYPECTQALYKCRFGYRSLGVNINVCCKGEMVALNPLR 78
QY 61 ICKRRCGHPGDTFFGSSFLAVGSEFEFGAKVYTCDEGYOLLGTEIDYRCADAGWTNDI 120
DB 79 KCKRRCGHPGDTFFGSSFLAVGSEFEFGAKVYTCDEGYOLLGTEIDYRCADAGWTNDI 138
QY 121 PICEVVKCLPYTELENGRITVSGAEPDOEYFGOVAFECNSFKITGCKEMHCSEGLW 180
DB 139 PICEVVKCLPYTELENGRITVSGAEPDOEYFGOVAFECNSFKITGCKEMHCSEGLW 198
QY 181 SNEKPCQVEISCLPPEVNGDGIYLRVYKENERFOYKCKGKRGDAVCTGSGMNP 240
DB 199 SKRKPCKVEISCLPPEVNGDGIYLRVYKENERFOYKCKGKRGDAVCTGSGMNP 258
QY 241 QPSCEMTCLTPYPIGNYITPHRIKIRIDELRECKNGFYPATRSPVSKCTTGWIPAP 300
DB 259 LPSCEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGTACTSTGWIIPAP 318
QY 301 RSLKRCDFPQFQHGHLTYEESRRYPVPIGKEYSYCDNGFTYSQSWDLRCVTNG 360
DB 319 RCLTKRCDFPQFQHGHLTYEESRRYPVPIGKEYSYCDNGFTYSQSWDLRCVTNG 378
QY 361 WEPEVCLFQCLFHYVEYESSYWMORRYIEGOSAKYQCHSGSYLSPNGQDTTYCTENGWSP 420
DB 379 WSPAVVCLFQCLFHYVEYESSYWMORRYIEGOSAKYQCHSGSYLSPNGQDTTYCTENGWSP 438
QY 421 PPKCVRIK 428
DB 439 TPKCIRVK 446

RESULT 3
CABP_HUMAN STANDARD; PRT; 597 AA.
AC P04003;
DT 23-OCT-1986 (rel. 02, Created)
DT 01-OCT-1993 (rel. 27, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE CAB-BINDING PROTEIN ALPHA CHAIN PRECURSOR (PROLINE-RICH PROTEIN)
DE (PRT).
GN CABPA OR CABP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RA MEDLINE: 90073699.
RA Matsuguchi T., Okamura S., Aso T., Sata T., Niho Y.;

RT RT "Molecular cloning of the cDNA coding for proline-rich protein (PRP):
RT RT identity of PRP as C4b-binding protein.";
RT RL Biochem. Biophys. Res. Commun. 165:138-144(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91113199.
RA Aso T., Okamura S., Matsuguchi T., Sakamoto N., Sata T., Niho Y.;
RT "Genomic organization of the alpha chain of the human C4b-binding
RT RT protein gene.";
RN [3]
RP Biochem. Biophys. Res. Commun. 174:222-227(1991).
RN [4]
RP SEQUENCE OF 9-81 FROM N.A.
RX MEDLINE: 88242821.
RA Linton S.J., Lewin A.R., Reid K.B.M.;
RT "Derivation of the sequence of the signal peptide in human
RT RT C4b-binding protein and interspecies cross-hybridisation of the C4bp
RT RT cDNA sequence.";
RN [5]
RP FEBS Lett. 232:328-332(1988).
RN [6]
RP SEQUENCE OF 203-288 FROM N.A.
RX MEDLINE: 86301119.
RA Linton S.J., Reid K.B.M.;
RT "Studies on the structure of the human C4b-binding protein gene.";
RN [7]
RP FEBS Lett. 204:77-81(1986).
RN [8]
RP SEQUENCE OF 80-597 FROM N.A.
RX MEDLINE: 86025405.
RA Chung L.P., Bentley D.R., Reid K.B.M.;
RT "Molecular cloning and characterization of the cDNA coding for C4b-
RT RT binding protein, a regulatory protein of the classical pathway of the
RT RT human complement system.";
RN [9]
RP Biochem. J. 230:133-141(1985).
RN [10]
RP SEQUENCE OF 49-88.
RX MEDLINE: 85296001.
RA Chung L.P., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence, studies of human C4b-binding protein: N-terminal
RT RT sequence analysis and alignment of the fragments produced by limited
RT RT proteolysis with chymotrypsin and the peptides produced by cyanogen
RT RT bromide treatment.";
RN [11]
RP Mol. Immunol. 22:427-435(1985).
RN [12]
RP ELECTRON MICROSCOPY, 3-DIMENSIONAL STRUCTURE, AND LIGAND BINDING.
RX MEDLINE: 83221615.
RA Dahlback B., Smith C.A., Mueller-Eberhard H.J.;
RT "Visualization of human C4b-binding protein and its complexes with
RT RT vitamin K-dependent protein S and complement protein C4b.";
RN [13]
RP Proc. Natl. Acad. Sci. U.S.A. 80:3461-3465(1983).
RN [14]
RP FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
RN [15]
RP ACTIVATION. IT BINDS AS A COFACTOR TO C3b/C4b INACTIVATOR
RN [16]
RP (C3bINA), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4b. IT
RN [17]
RP ALSO ACCELERATES THE DEGRADATION OF THE C4bC2a COMPLEX (C3
RN [18]
RP CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2a. ALPHA
RN [19]
RP CHAIN BINDS C4b. IT INTERACTS ALSO WITH ANTICOAGULANT PROTEIN S
RN [20]
RP AND WITH SERUM AMYLOID P COMPONENT.
RN [21]
RP SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
RN [22]
RP OF 3 POSSIBLE SORTS: A 570 KDA COMPLEX OF 7 ALPHA CHAINS AND 1
RN [23]
RP BETA CHAIN, A 530 KDA HOMOPENTAMER OF ALPHA CHAINS OR A 500 KDA
RN [24]
RP COMPLEX OF 6 ALPHA CHAINS AND 1 BETA CHAIN. THE CENTRAL BODY OF
RN [25]
RP THE ALPHA CHAIN HOMOPOLYMER SUPPORTS TENTACLES, EACH WITH THE
RN [26]
RP BINDING SITE FOR C4b AT THE END.
RN [27]
RP TISSUE SPECIFICITY: CHYLOMICRONS IN THE PLASMA.
RN [28]
RP 1- SIMILARITY: CONTAINS 8 SUSHI (SCR) REPEATS.
RN [29]
RP 1- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOBOLIPROTEIN R.
RN [30]
RP 1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.
RN [31]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN [32]
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN [33]
RP the European Bioinformatics Institute. There are no restrictions on its
RN [34]
RP use by non-profit institutions as long as its content is in no way
RN [35]
RP modified and this statement is not removed. Usage by and for commercial
RN [36]
RP entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN [37]
RP or send an email to license@sib-sib.ch).
```



```

CC -----
DR EMBL: M31452: AAA36507.1: -
DR EMBL: M62486: AAA36506.1: -
DR EMBL: M62475: AAA36506.1: JOINED.
DR EMBL: M62476: AAA36506.1: JOINED.
DR EMBL: M62477: AAA36506.1: JOINED.
DR EMBL: M62478: AAA36506.1: JOINED.
DR EMBL: M62479: AAA36506.1: JOINED.
DR EMBL: M62480: AAA36506.1: JOINED.
DR EMBL: M62481: AAA36506.1: JOINED.
DR EMBL: M62482: AAA36506.1: JOINED.
DR EMBL: M62484: AAA36506.1: JOINED.
DR EMBL: M62485: AAA36506.1: JOINED.
DR EMBL: M62486: AAA36506.1: JOINED.
DR EMBL: X07853: CA330701.1: -
DR EMBL: X04284: CAB31244.1: -
DR EMBL: X04296: CA27839.1: -
DR EMBL: X02865: CA26617.1: -
DR PIR: A3368: NBHUC4.
DR HSSP: P10998: LYVC.
DR MIM: 120830: -
DR INTERPRO: IPR000436: -
DR PRAM: PF00084: sushi; 8.
KW Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal;
KW Polymorphism.
FT SIGNAL 1 48
FT CHAIN 49 597 C4B-BINDING PROTEIN ALPHA CHAIN.
FT DOMAIN 49 539 8 X SUSHI (SCR) REPEATS.
FT REPEAT 49 109 SUSHI 1.
FT REPEAT 112 171 SUSHI 2.
FT REPEAT 174 235 SUSHI 3.
FT REPEAT 238 295 SUSHI 4.
FT REPEAT 298 361 SUSHI 5.
FT REPEAT 364 423 SUSHI 6.
FT REPEAT 425 481 SUSHI 7.
FT REPEAT 483 539 SUSHI 8.
FT DISULFID 50 96 BY SIMILARITY.
FT DISULFID 81 108 BY SIMILARITY.
FT DISULFID 113 154 BY SIMILARITY.
FT DISULFID 140 170 BY SIMILARITY.
FT DISULFID 175 217 BY SIMILARITY.
FT DISULFID 203 234 BY SIMILARITY.
FT DISULFID 239 281 BY SIMILARITY.
FT DISULFID 297 294 BY SIMILARITY.
FT DISULFID 299 348 BY SIMILARITY.
FT DISULFID 332 360 BY SIMILARITY.
FT DISULFID 364 387 BY SIMILARITY.
FT DISULFID 7365 7409 BY SIMILARITY.
FT DISULFID 7399 422 BY SIMILARITY.
FT DISULFID 426 468 BY SIMILARITY.
FT DISULFID 454 480 BY SIMILARITY.
FT DISULFID 484 525 BY SIMILARITY.
FT DISULFID 511 538 BY SIMILARITY.
FT DISULFID 546 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT DISULFID 558 558 INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .).
FT VARIANT 92
FT VARIANT 357
FT SEQUENCE 597 AA: 67033 MW: 67E03FEA85A16DD CRC64:

```

```

DB 107 -FCIYKRRHRELRLNG--QVEIKTDLSPGSOIESSCSGFFLLISTSTR-CEVDDRGV 162
QY 116 WTNDIPICEVVKCLPVTLENGRIYSGAEPDQVEYFGQVREECNSGKIEGQEMHCS 175
DB 163 WSHPLPOCEIYVCKRPPIIDRNG---HSGEENFYAGSHVYISCDPPRSLIGHASISCT 218
QY 176 -EN---GLMSNKKPOCVLEISCLPRVENDGITY-LKPYKKEHERQYCKQGFYKKEKD 230
DB 219 VENETIGWRPSPPTCEKITCKRPVSHGEMVSGCPILYNVDITVFCQKGFVL--KCS 276
QY 231 AV--C-TSGWMPQ-PSCSEMTCLF-PYIPNGIY--PHRIK---HRIDETRYCKNKF 280
DB 277 SYIHCDAASKNPPSPACEPNCINLPDIIPASMETYPRPTEDYVVGTVLRYKCHRY 336
QY 281 YPATRSPVSKCTITG--WIPAPRCILKPDPPQFHKGLRYEESRRP----- 325
DB 337 KPTDEPTTVICQKMLRMTPYGCEALCPPEKLNNGEITQHRKSRPANHCYFYGDEIS 396
QY 326 -----YPPVPIG-----KEYSYCDN 341
DB 397 FSCHETSRESALCQDGTWSPPTPSCGDCINPPRIAHGNYKOSSSYSEKKEIIECDK 456
QY 342 GPTPSOSYWDYLRCTVNGMEVEVP-CLRQCIFHYVEGESSYWMORRYEGQSAKQCHS 400
DB 457 GYIILVGA---KLSCSYMSAPAPQCKALCKRPVELNGLRSLVDQYEPENVTIQDS 513
QY 401 GYSLPNGDYYCTEN-GWSP-PRC 424
DB 514 GYGAV-GPQSTICSGNRTWYPEVPC 538

```

```

RESULT 4
C4BP_BOVIN STANDARD: PRT; 610 AA.
ID C4BP_BOVIN
AC 028065;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE C4B-BINDING PROTEIN ALPHA CHAIN PRECURSOR.
OS C4BPA.
OS Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Bovinae; Bos.
[1]
SEQUENCE FROM N.A.
RP TISSUE=LIVER;
RC MEDLINE: 95015909.
RX Hillarp A., Thern A., Dahlbaeck B.:
RT "Bovine C4b binding protein. Molecular cloning of the alpha- and
RT beta-chains provides structural background for lack of complex
RT formation with protein S."
RL J. Immunol. 153:4190-4199(1994).
CC -!- FUNCTION: C4BP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT
CC ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR
CC (C3BIN), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT
CC ALSO ACCELERATES THE DEGRADATION OF THE C4B2A COMPLEX (C3
CC CONVERTASE) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA
CC CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYLOID P COMPONENT.
CC -!- SUBUNIT: DISULFIDE-LINKED COMPLEX OF C4BP ALPHA AND BETA CHAINS
CC (BY SIMILARITY).
CC -!- SIMILARITY: TO C4BP BETA CHAIN AND TO PIG APOLIPOPROTEIN R.
CC -!- SIMILARITY: CONTRAINS 8 SUSHI (SCR) REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



FT	DISULEFID	2328	2351		BY SIMILARITY.
FT	DISULEFID	2329	2373		BY SIMILARITY.
FT	DISULEFID	2363	386		BY SIMILARITY.
FT	DISULEFID	390	431		BY SIMILARITY.
FT	DISULEFID	417	443		BY SIMILARITY.
FT	DISULEFID	447	488		BY SIMILARITY.
FT	DISULEFID	474	501		BY SIMILARITY.
FT	DISULEFID	509	509		INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	DISULEFID	521	521		INTERCHAIN (WITH BETA CHAIN) (POTENTIAL).
FT	CAROXYHD	31	31		N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CAROXYHD	177	177		N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CAROXYHD	186	186		N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CAROXYHD	459	469		N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CAROXYHD	491	491		N-LINKED (GLCNAC. . . ) (POTENTIAL).
SQ	SEQUENCE	558 AA;	62266 MW;	592FC0667ED1E5FF CRC64;	

Query Match
18.3%; Score 449; DB 1; Length 558;

Best Local Similarity
26.7%; Pred. No. 4,6e-27;

Matches 137; Conservative
71; Mismatches 198; Indels 108; Gaps 24;

QY	5	GPPPRNSIIILGSSNEQLSECTQATYYCRGY -RTLGTIVAKCKN-GEMVPSNDRIC	62
DB	16	GPDPALPALASEMNOJDFESHHTLLRLRYCRGYSRASSOSLCPDLKRW--QINIK	72
QY	63	RRRPGHPDTPFGSFRILWGSFEFGAKVVNYDCGYOLLG-EIDYRECDAG--WTND	119
DB	73	VAKSCRNPDDLONG-KVEVKTDLFLEGSIERSCSGIYLISSTSYCIQKGYSWSDP	130
QY	120	IPICEVYAKLPYTELENGRHIVSGAAEPDOEYFEGGVAFRECNSGFRIEQKEMHS---	175
DB	131	LPECVIAKCGMPDINSNGR---HGREGEEFFTYRSSVYKCDPDPTLLGNASITCTVANK	187
QY	176	ENGLSMNEKPQCFVETLSCLPRVENGD-GIYLKPVKKENRFPYKKQGGVYKERDAVCT	234
DB	188	TYGVWSPSPPTCERLIICPMFKVLHGINSGEFHRTKYKDSVFRCQKGVNLGSGVTHCE	247
QY	235	GSG-WNOPSCSEMCL-TPLYIPNG-IYTPHRIR---HRIDEIRYEKNGFYPATRSP	287
DB	248	ADGMSWPVPVELNSTCIDPIDIPNALITSPPRPREDYVPQVTIARIICRPTEPATRP	307
QY	268	VSKCI-----TGMLPARPCSLKPCDFQPFKIGRLYIESRRP-----YFPVPIKEYSY	337
DB	308	M---TVICQKDLWSMLROCKEICCPVPDPKSVRIQHKAHPDMDCIFY--FGDEVSY	361
QY	338	YCDN-----GPTFPSOZYMYLR-----	355
DB	362	TCONIMLTATCKSDGTWHPRTPSCHOSCDFPAALAHGHTKSSSY-YVRQVYEEBE	419
QY	356	-----CTVGMGEPEVP-CLROCI FHYVEYGESSYMORRYIEGOSAKVOCBSGYS	403
DB	420	GYLIVEATITSCWYSQMTPAAPQCALRKRPDIGNVGLTNKDQVETENVTIQCDSGFV	479
QY	404	LPNODDYTCENG-NSP-----PPKCARI	427
DB	480	M-LGSQSITCSENGTYPRKVSRCDEVEPRDCEHV	512

RESULT 7

CRL\_HUMAN STANDARD: PRT: 2039 AA.

AC PI7927,

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE COMPLEMENT RECEPTOR TYPE 1 PRECURSOR (CB3/4B RECEPTOR) (CD35 ANTIGEN)

GN CRI OR C3BR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NN [1]

PP SEQUENCE FROM N.A.

[illegible]

Query Match	Best Local Similarity	Score	DB 1	Length	2039
Matches 132	Conservative 63	Mismatches 149	Indels 132	Gaps 31	
QY 27 GQATATYKCRPGRTLTGTYKVKCKNGEWPVNSNRSLKCRPGHGGDTPGSRFLAGSEF 86	11 :   :	11 :   :	11 :   :	11 :   :	
DB 66 GTTYLNECRPGTSGRPFSLTCLKNVSWTGAKDR--CRKSCSNRPDPVPAWGVHYIKG--I 121	:	:	:	:	
QY 87 BEGAKVVTYCEGYOLLEIDRECDADG---WTNDIPICEVVK-LPTELENGRIYS 141	:   :	:   :	:   :	:   :	
DB 122 QGSGQIKYSCTGRTGRLGS-SSATCIISGDTYIMDEPDCRIPGCLPPT--ITNDGFS 179	:   :	:   :	:   :	:   :	
QY 142 GAAPDQETVRCQVVRFECSNG-----FKIGQKEMHCSN-----GLMSNEKPC-VKI 190	:	:	:	:	
DB 180 TNRE---NFHYGSVYTYRCPNPGSGGRKVFELGEPISITCTSDDDQVIGMSGAPQCIIN 236	:   :	:   :	:   :	:   :	
QY 191 SCLPVRVENG-----DGITYLKPVYKKEENERQYCKQGGFYKKEGDAVCTG-SGWNPO-PSC 244	:	:	:	:	
DB 237 KCTPRNVEGILVSDN---RSLFSLNVEYEFRCQGFYVWKGRPRYKCOALNKRPEPLPSC 293	:   :	:   :	:   :	:   :	
QY 245 EEMTCLTPYIPNGIYTPPHRIKIRID---DETRECKNGFYPATRSPVSKCTYTG-WID- 298	:   :	:   :	:   :	:   :	
DB 294 SHV-QQPP--PVLHAHERQBDKDNFSPQGEVIFYSEPG-YDLRGAASNRCPPOGWSMA 349	:   :	:   :	:   :	:   :	
QY 299 APRSLKPCD--FPQFKHGLRYEESRRPFPV--PIKEIYTYCDNGTGT--PQSOTYD 352	:	:	:	:	
DB 350 APTCEVKSCDDEMGQLMNRV-----LFPVNIQLGAKVYVCEDEGQFOLKGGSSASY-- 399	:   :	:   :	:   :	:   :	
QY 353 YLRCTVNG---WEPEVPLRQCFIHYVYEGSSYQWRATYIGQSAKVOCHSYSLP-N- 406	:   :	:   :	:   :	:   :	
DB 400 ---CYLAGHESIAMNSNVPCED-----ITCPSPVPVLPNKR 431	:   :	:   :	:   :	:   :	

OY	407		-----GOOYYCTC-----NG-W-SPPC	424
			I:	I
			:	:
Db	432	HTEKLEFVPEPKAVNYTCDDPHDGTSTDLIGESTICTSDPGNGWSSPAPKC		487
RESULT	8			
CAB_MOUSE				
ID	CAB_MOUSE	STANDARD:	PRT:	469 AA.
AC	P06607,			
DT	01-AUG-1988 (Rel. 08, Created)			
Dt	01-FEB-1994 (Rel. 28, Last sequence update)			
Dr	01-MOV-1997 (Rel. 35, Last annotation update)			
DE	CAB-BINDING PROTEIN PRECURSOR (CABP).			
GN	CABPA OR CABP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrate; Euteleostomi;			
RN	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
RP	[1]			
RX	SEQUENCE FROM N.A. MEDLINE: 88024997.			
RA	Kristensen T., Ogata R.T., Chung L.P., Reid K.B.M., Taek B.F.; "cDNA structure of murine Cab-binding protein, a regulatory component of the serum complement system."; Biochemistry 26:1668-1674(1987)."			
RL	Biochemistry 26:1668-1674(1987).			
CC	-I FUNCTION: CABP CONTROLS THE CLASSICAL PATHWAY OF COMPLEMENT ACTIVATION. IT BINDS AS A COFACTOR TO C3B/C4B INACTIVATOR (C3BIN), WHICH THEN HYDROLYZES THE COMPLEMENT FRAGMENT C4B. IT ALSO ACCELERATES THE DEGRADATION OF THE C4BC2A COMPLEX (C3 CONVERTAS) BY DISSOCIATING THE COMPLEMENT FRAGMENT C2A. ALPHA CHAIN BINDS C4B. IT INTERACTS ALSO WITH SERUM AMYOLOID P COMPONENT. -I SUBUNIT: HOMOPOLYMER; NOT COVALENTLY LINKED. MOUSE LACKS THE BETA CHAIN OF CABP.			
CC	-I SIMILARITY: TO CABP BETA CHAIN AND TO BIG A POLIPOPROTEIN R.			
CC	-I SIMILARITY: CONTRAINS 6 SUSHI (SCR) REPEATS.			
CC	-I CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-44 IS THE INITIATOR.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial/ entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M17122; AAA37312.1; ALT_INIT.			
DR	PIR: A27117; NBMSC4.			
DR	HSSP: P10998; LYVC.			
DR	GMD: MG1:88229; CABP.			
DR	INTERPRO: IPRO00436; "			
KM	PFAM: PF00084; sushi; 6.			
FT	SIGNAL	1	56	
FT	CHAIN	57	469	CAB-BINDING PROTEIN.
FT	DOMAIN	57	414	6 X SUSHI (SCR) REPEATS.
FT	REPEAT	57	116	SUSHI 1.
FT	REPEAT	119	177	SUSHI 2.
FT	REPEAT	180	241	SUSHI 3.
FT	REPEAT	244	300	SUSHI 4.
FT	REPEAT	302	356	SUSHI 5.
FT	REPEAT	358	414	SUSHI 6.
FT	DISULFID	58	103	BY SIMILARITY.
FT	DISULFID	88	115	BY SIMILARITY.
FT	DISULFID	120	160	BY SIMILARITY.
FT	DISULFID	146	176	BY SIMILARITY.
FT	DISULFID	181	223	BY SIMILARITY.
FT	DISULFID	209	240	BY SIMILARITY.
FT	DISULFID	245	287	BY SIMILARITY.
FT	DISULFID	273	299	BY SIMILARITY.
FT	DISULFID	303	343	BY SIMILARITY.
FT	DISULFID	329	355	BY SIMILARITY.
FT	DISULFID	359	400	BY SIMILARITY.
FT	DISULFID	386	413	BY SIMILARITY.



```

FT DISULFID 431 458 BY SIMILARITY.
FT DISULFID 463 501 BY SIMILARITY.
FT DISULFID 487 514 BY SIMILARITY.
FT DISULFID 519 568 BY SIMILARITY.
FT DISULFID 548 585 BY SIMILARITY.
FT DISULFID 594 636 BY SIMILARITY.
FT DISULFID 622 649 BY SIMILARITY.
FT DISULFID 654 689 BY SIMILARITY.
FT DISULFID 675 704 BY SIMILARITY.
FT DISULFID 709 752 BY SIMILARITY.
FT DISULFID 738 769 BY SIMILARITY.
FT DISULFID 778 820 BY SIMILARITY.
FT DISULFID 806 833 BY SIMILARITY.
FT DISULFID 841 884 BY SIMILARITY.
FT DISULFID 870 897 BY SIMILARITY.
FT DISULFID 902 945 BY SIMILARITY.
FT DISULFID 931 958 BY SIMILARITY.
FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 813 813 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 989 921 YGS -> EFR (IN REF. 4).
FT CONFLICT 306 306 S -> T (IN REF. 2).
FT CONFLICT 520 520 P -> A (IN REF. 2).
FT CONFLICT 962 963 MISSING (IN REF. 4).
SO SEQUENCE 1025 AA; 112994 MW; 19E518B9A0273094 CRC64;

```

Query Match 15.2%; Score 373; DB 1; Length 1025;  
 Best Local Similarity 25.7%; Pred. No. 5.5e-21;  
 Matches 139; Conservative 61; Mismatches 184; Indels 156; Gaps 35;

```

OY 3 CKGPP-----PRENSEISGKSEQLYSEGTQATYKCRPGYTLCTYKVC-KNGEWPS 56
DB 402 COAPKIKINGOKEDSYLLN-----FPDGTISIRYSCDPGYLLVGEDTHTCHPECKMPPI 454
OY 57 NPS-RICKRRPGC-HPGDTPPGSF-RLAVGSFEFGAKVYVTCDCGYLLGIDIRECA 113
DB 455 TPQCTVAGCKPVPGLFRPQNOFTRTAVNS-----SCDGGFO-LSBSAVQLCGG 503
OY 114 D-GWNTNDIPICEVVKCLPVTLENGRIYSGAEPDOEYFGGVFECSNG-----FKI 166
DB 504 TTPMTEIRLCKEITCPPPVYIHNGTHWSSSE---DVPYGVIVYVYMCYRPGEGEGRFL 560
OY 167 EGQKMHCSN-----GLMSNEKPGC---VETISCLPPRVENGDTYL--KPYKENERF 215
DB 561 IGEQTHCTSDSRGRGSSWSPAPLKLSPAVQCTDVHEN--GVKLTDNKAPAFYNDVS 618
OY 216 OYKCKGQVYKERNDAVCTGSG-WMPQ-----PSCDEN-----TC--- 249
DB 619 MFKCDGDTLSSGSOIRCKANNTWDEPKLCKEKGCEPERVHGLDPDSDHILVKRFGONG 678
OY 250 --LT-----PYIPGICITTPRIKIRI--DDEIRYEC 276
DB 679 YQLTGVYKCKQNAENGWTFMKIEVCTVILCOPPKINAGHTGMAKHFLIGNESVYSC 738
OY 277 KNGFPAIRSPYSKCTIT-----GWI-PAPRC---SLKPCFPPOKRGRLYEESSRPY 326
DB 739 DGGFYLAGEKSL-QCVNDSKSGHSWSPPOCLOSSPLHTCHPDPEYKHG----- 786

```

```

OY 327 FVPVIGKEYS-----YYCDNGFTTPSQSYWDYLRG-TYNGWPEYEP-CLRCIOF-- 373
DB 787 --YKLKTHASASHNDIVFVCHNOGFIANGS---HLIRCHTNNTVLPVPTCIRKASLCC 841
OY 374 ---HYVEGESSYWO-RRYIEGOSAKVOCCHSGYSLPFGODTYCTENG--WS-PPPKCVRI 427
DB 842 QSPSTIPNHNHGGSLARPPGMSVMSYSCYGF-LMAGEARLICTHEGTVMSQPPPCREYV 900

RESULT 10
CR2_HUMAN
ID CR2_HUMAN STANDARD; PRT; 1033 AA.
AC P20023;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR)
DE (EPSTEIN-BARR VIRUS RECEPTOR) (CD21 ANTIGEN).
GN CR2 OR C3DR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominoidea; Homo.
OC [1]
RP SEQUENCE FROM N.A.
RA Fujisaku A., Harley J.B., Frank M.B., Gruner B.A., Frazier B.,
RA Holers V.M.;
RA "Genomic organization and polymorphisms of the human C3d/Epstein-Barr
RT virus receptor.";
RL J. Biol. Chem. 264:2118-2125(1989).
[2]
RP SEQUENCE OF 226-233; 256-267; 332-341; 667-677 AND 898-908.
RA MEDLINE: 86287311.
RA Weis J.J., Fearon D.T., Klickstein L.B., Wong W.W., Richards S.A.,
RA de Bruyn Kops A., Smith J.A., Weis J.H.;
RT "Identification of a partial cDNA clone for the C3d/Epstein-Barr
RT virus receptor of human B lymphocytes: homology to the receptor for
RT fragments C3b and C4b of the third and fourth components of
RT complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5639-5643(1986).
CC -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D AND FOR THE EPSTEIN-BARR
CC VIRUS ON HUMAN B-CELLS. PARTICIPATES IN B LYMPHOCYTES ACTIVATION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MATURE B LYMPHOCYTES AND FOLLICULAR
CC DENDRITIC CELLS OF THE SPLEEN.
CC -!- SIMILARITY: TO MOUSE CR2. CD21 IS A MEMBER OF THE REGULATORS OF
CC COMPLEMENT ACTIVATION (RCA) FAMILY.
CC -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) REPEATS.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD21 entry;
CC WWW="HTTP://WWW.NCBI.NLM.NIH.GOV/PROW/CD/CD21.HTM".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DB EMBL: M26004; AAA35786.1; -.
DB EMBL: M26016; AAB04638.1; -.
DB EMBL: M24007; AAB04638.1; JOINED.
DB EMBL: M24008; AAB04638.1; JOINED.
DB EMBL: M24009; AAB04638.1; JOINED.
DB EMBL: M24010; AAB04638.1; JOINED.
DB EMBL: M24011; AAB04638.1; JOINED.
DB EMBL: M24012; AAB04638.1; JOINED.
DB EMBL: M26009; AAB04638.1; JOINED.
DB EMBL: M26010; AAB04638.1; JOINED.
DB EMBL: M26011; AAB04638.1; JOINED.
DB EMBL: M26012; AAB04638.1; JOINED.
DB EMBL: M26013; AAB04638.1; JOINED.
DB EMBL: M26014; AAB04638.1; JOINED.
DB EMBL: M26015; AAB04638.1; JOINED.

```



```

DR PIR: A32036; A32036.
DR PIR: A24319; A24319.
DR PIR: B24319; B24319.
DR PIR: C24319; C24319.
DR PIR: D24319; D24319.
DR PIR: E24319; E24319.
DR PIR: F24319; F24319.
DR HSP: P10998; 1VVC.
DR MIM: 120650; -.
DR INTERPRO: IPR00436; -.
DR PFM: PFM0084; sushi: 15.
KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
  Receptor; Sushi.
FT SIGNAL 1 20
FT CHAIN 21 1033
FT DOMAIN 21 971
FT TRANSMEM 972 999
FT DOMAIN 1000 1033
FT DOMAIN 22 971
FT REPEAT 22 83
FT REPEAT 90 147
FT REPEAT 153 211
FT REPEAT 214 272
FT REPEAT 275 343
FT REPEAT 350 407
FT REPEAT 409 467
FT REPEAT 470 523
FT REPEAT 526 594
FT REPEAT 601 658
FT REPEAT 715 715
FT REPEAT 718 780
FT REPEAT 787 844
FT REPEAT 850 908
FT REPEAT 911 969
FT DISULFID 23 65
FT DISULFID 51 82
FT DISULFID 91 132
FT DISULFID 116 146
FT DISULFID 154 197
FT DISULFID 183 210
FT DISULFID 215 256
FT DISULFID 242 271
FT DISULFID 276 325
FT DISULFID 305 342
FT DISULFID 351 393
FT DISULFID 379 406
FT DISULFID 410 453
FT DISULFID 439 466
FT DISULFID 471 509
FT DISULFID 495 522
FT DISULFID 527 576
FT DISULFID 556 593
FT DISULFID 602 644
FT DISULFID 630 657
FT DISULFID 662 699
FT DISULFID 685 714
FT DISULFID 719 762
FT DISULFID 748 779
FT DISULFID 788 830
FT DISULFID 816 843
FT DISULFID 851 894
FT DISULFID 880 925
FT DISULFID 912 955
FT DISULFID 941 968
FT CARBOHYD 121 121
FT CARBOHYD 127 127
FT CARBOHYD 294 294
FT CARBOHYD 372 372
FT CARBOHYD 492 492
FT CARBOHYD 623 623
FT CARBOHYD 682 682
FT CARBOHYD 800 800
FT CARBOHYD 823 823
  
```

```

FT CARBOHYD 861 861
FT CARBOHYD 911 911
FT CONFLICT 667 667
FT CONFLICT 902 902
FT CONFLICT 906 906
SQ SEQUENCE 1033 AA; 112973 MW; 1749DBA407847ADA CRC64;

Query Match 15.0%; Score 368; DB 1; Length 1033;
Best Local Similarity 24.3%; Pred. No. 1,3e-20;
Matches 131; Conservative 63; Mismatches 198; Indels 146; Gaps 31;

QY 1 EDCKGPPRENSFILLSSWSEQ--LYSEGTQATYKCRPGYRTLG-TIVKCKNGEWPS 56
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 KECQAP-----NILNQKEDRHVFRDPGTSIKYCNPGVILGERSIQTSSEGVWTP 462

QY 57 NPSRLCKRRPGHRRDPFGFSRLAVGSEEF-CAKVYVCDESYOLGELDYRECAD- 114
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 463 VPQ--CKVAACEATGR-----QLTRPQHFNRPDYNSSCGEGYKLSGV-YQECGCTI 513

QY 115 GWTNDIPICEVVKCLPYTELENGRIVSGAAEPDQEVYFGVYRPECNSG-----FKIEG 168
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 PMWEIRLCKEITICPPRPVYINGAHGSSLE--DFYGTIVYTCNPGDERGVEFSLIG 570

QY 169 QKEMHCSEN---GLWSNEKPOC---VETISCLPPRVENGDIYLPV-YKENERFOYKC 219
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 ESTIRCTSNQERGTWSGPAFLKLSLAVOCSHVHIANGYKISGRAPFYNDYTFKC 630

QY 220 KQGVYKERDVAOCVTSGC--NNPO--PSCSEMTC----- 249
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 631 YSGFTLKGSSQIRCKADNDWDELPVCEKETCOHVROSLOELPAGSVEYLVNMSCDQGY 690

QY 250 LT-----PYINGIYTPHRIKRI--DDEIRYECKN 278
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 LTGHAYQMCDAENGIMFKKIPLOCKVICHPPVYINGKHTGMMAENFLGNEVSYECDQ 750

QY 279 GFYPATSPYSKCTIT-----GWT-PAPRCSLRP-----CPEPQFKHRLYYEESRRPYFP 328
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 751 GFYLLGERKL-QCSDSKSGHSGWSGSPQCLRSPVTRCPNPEVKHG----- 796

QY 329 VPIKEYSY-----CDNGFTTPSOSYWDLR--TYNGMPEPEP--CLROCFHYVE 377
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 797 YKLNKTHSAISHNDIYVDNCPGIMNGSR---VIRCHTDNTWYPGVPTCIKAFICPP 853

QY 378 YGESSYWQR-----RYIESQSAKVVOCHSGSYSLPNQDYYCTENG--WS--PPKCVRI 427
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 854 PPKTPNGNHTGNIARFSPGMSILYSCDGYLLV-GEALLICTHEGTWSQPAFHCKEV 910

RESULT 11
F13B_HUMAN STANDARD; PRT; 661 AA.
ID F13B_HUMAN AC P05160;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COAGULATION FACTOR XII B CHAIN PRECURSOR (EC 2.3.2.13) (PROTEIN-
DE GLUTAMINE GAMMA-GLOTTAMYLTRANSFERASE B CHAIN) (TRANSGLOTTAMINASE B
DE CHAIN).
DE F13B.
GN Homo sapiens (human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91105054.
RA Botteus R.E., Ichinose A., Davie E.W.;
RT "Nucleotide sequence of the gene for the b subunit of human factor
  XII ";
RL Biochemistry 29:11195-11209(1990).
RN [2]
RP SEQUENCE OF 2-661 FROM N.A.
RX MEDLINE: 87026535.
  
```



Query	1	EDCKPPPRENSSEILSSGWS--EOLYSGCATYKCRPGYTTLD---TIYVCKNGBM--	53
Db	151	ETCLAP-----ELYNGYSTQTQTFKFKKQVQYECATGYTTAGSKTEVECLLTYSWL	204
Qy	54	VPSNDRICRRKPCGHPEDTPFGSFRL-----AVGSEFEQKAVYVTCDEGYLLGEI	106
Db	205	TPK-----CTKLKC-----SSLRLIENGYFHPYKQYEEGDVVOFPCHENYISGS-	250
Qy	107	DYRECDADGWINDPICE--VVKCLPTLELNGITVSGAEPDEYFFQGVNPECSGF	164
Db	251	DLIOCYNGWVPESPVECDGRNRCPPPLPINSKIQYHST---TYHGEIVHIECLNF	306
Qy	165	KIEGQKMHCSGSLMSNEKQVVE---ISC-LPPRYENGDIYLYLVYKENEREQYKC	219
Db	307	EIHSGAEIIRC-EDKKWT-EPKRCIEGQEKVACEEPTEINGAANLHSKIYNGDKTYAC	364
Qy	220	KQGFYKRGAGVAVCTGSGWNPQSCEE--MTCL-TYIIPGICYTPHPI-KHRIDDELRYE	275
Db	365	KSGYLLHGSNETITNRCWKTLPPRCVLENENCKHPVVMGAADGILASVATGSSVEYR	424
Qy	276	CKNGEYPATRSVSKCTTTGMPAPRCSLKPCHDPQFQKRLHYEESRRPYFPVPIKEY	335
Db	425	C-NEYLLRGSKISRCEQKMSPPVC-LEPCYVNVDMNRNNIMKWKYEGVLAHDLI	482
Qy	336	SYCDNETTFSQSYWYDLKCTVNGMPEVP-CLRQ-----CIFHYVEGESSTW	384
Db	483	DFVCKQGDLSPLPLSELVQVCRGGEVKKYPLCTRKESKGMCTSPPLIKHGVIIISF---	539
Qy	385	QRRYIEGSAVQCHSGSYLPGNGDITYYCENGMSPPCKV	425
Db	540	VDIYENGSSVYERCFDHFHLEGSREA-YCLDGMWTTPLCL	579

RA Nonaka M., Matsuda Y., Shirolshi T., Moriaki K., Nonaka M.,  
 RA Natsune-Sakai S.:  
 RT "Molecular cloning of the b subunit of mouse coagulation factor XIII  
 RT and assignment of the gene to chromosome 1: close evolutionary  
 RT relationship to complement factor H.",  
 RL Genomics 15:535-542(1993).  
 CC -1- FUNCTION: THE B CHAIN OF FACTOR XIII IS NOT CATALYTICALLY ACTIVE,  
 CC BUT IS THOUGHT TO STABILIZE THE A SUBUNIT'S AND REGULATE THE RATE  
 CC OF TRANSGLUTAMINASE FORMATION BY THROMBIN.  
 CC -1- SUBUNIT: TETRAMER OF TWO A CHAINS AND TWO B CHAINS.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LIVER AND KIDNEY.  
 CC -1- SIMILARITY: CONTAINS 10 SUSHI (SCR) REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D10071; BAA00963.1; -  
 DR PIR: A46013; A46013.  
 DR HSSP: P10998; 1VVC.  
 DR MGI: 88379; F13B.  
 DR INTERPRO: IPR000436; -  
 DR PFAM: PF00084; sush1; 8.  
 KW Transferrase; Plasma; Blood coagulation; Repeat; Glycoprotein; Signal;  
 KW Sush1.  
 FT SIGNAL 1 20 BY SIMILARITY  
 FT CHAIN 21 668 COAGULATION FACTOR XIII B CHAIN.  
 FT DOMAIN 24 647 10 X SUSHI (SCR) REPEATS.  
 FT REPEAT 24 88 SUSHI 1.  
 FT REPEAT 90 147 SUSHI 2.  
 FT REPEAT 152 209 SUSHI 3.  
 FT REPEAT 212 268 SUSHI 4.  
 FT REPEAT 273 328 SUSHI 5.  
 FT REPEAT 335 390 SUSHI 6.  
 FT REPEAT 395 451 SUSHI 7.  
 FT REPEAT 453 516 SUSHI 8.  
 FT REPEAT 523 579 SUSHI 9.  
 FT REPEAT 581 647 SUSHI 10.  
 FT DISULFID 25 76 BY SIMILARITY.  
 FT DISULFID 59 87 BY SIMILARITY.  
 FT DISULFID 91 135 BY SIMILARITY.  
 FT DISULFID 118 146 BY SIMILARITY.  
 FT DISULFID 153 197 BY SIMILARITY.  
 FT DISULFID 180 208 BY SIMILARITY.  
 FT DISULFID 213 225 BY SIMILARITY.  
 FT DISULFID 241 267 BY SIMILARITY.  
 FT DISULFID 274 316 BY SIMILARITY.  
 FT DISULFID 302 327 BY SIMILARITY.  
 FT DISULFID 336 378 BY SIMILARITY.  
 FT DISULFID 364 389 BY SIMILARITY.  
 FT DISULFID 396 439 BY SIMILARITY.  
 FT DISULFID 425 450 BY SIMILARITY.  
 FT DISULFID 454 505 BY SIMILARITY.  
 FT DISULFID 486 515 BY SIMILARITY.  
 FT DISULFID 524 567 BY SIMILARITY.  
 FT DISULFID 553 578 BY SIMILARITY.  
 FT DISULFID 582 636 BY SIMILARITY.  
 FT DISULFID 616 646 BY SIMILARITY.  
 FT CARBOHYD 162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 162 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 545  
 FT SEQUENCE 668 AA: 76078 MW: 80BC9E0DA9E53FA6 CRC64;  
 QY 13.58; Score 332; DB 1; Length 668;  
 Best Local Similarity 27.08; Pred. No. 4.5e-18;  
 Matches 129; Conservative 48; Mismatches 199; Indels 102; Gaps 29;  
 1 EDDKGPPEPNSLSSGMSBOLYSEGTQATYK-----CRGYRFL--GTIVKCK 49

DB 151 ETCLAP-----ELEGNST-----TQTFKVDIVATCTAGYTTTGKOTGEACQ 198  
 QY 50 NGEM--VPSNPSRICRRPCGHPDPFGSFL-----AVGSEFECAKVVYTDGEGY 100  
 DB 199 ANWSLTPQ-----CKRLMC-----SSLRLIENGYPHVPQYTEEDGLVQPCHEHY 245  
 QY 101 QLLGEIDYRCADGWTNDIPICE--VVKCLPYTELENGRIVSGAEP-DOERYFGQVVR 157  
 DB 246 YLSGS-DLIDCYNFGWPEESPICEGRNRCPPPVPLNSKI-----QPHSTYRHGERVH 299  
 QY 158 FECSGKIEGKEMHOSCEGWSNEKPOVE-----ISC-LPRVENGDSIYLKPYKEN 212  
 DB 300 IEBELNFVIGSSEELC-ENGKMT-EPKCTEKEKACQPSVEGVAHPSEIYISG 357  
 QY 213 ERFQYKQGFYKENDAVCTSGMNPQSCSE--MTCL-TPYIPNGIYTPHRI-KHRI 268  
 DB 358 DKYTRCGGGYSLRGSSTICNCRGRTLPPECYENIENCKAPPDPIANGVVDGLASYTT 417  
 QY 269 DDEIRYCKNGFYPATRSPVSKCTITGWTAPKCSLKPCCFP-----QPK----- 313  
 DB 418 GSSVEYKRC-NEYLLKSGSESRCEGAMSSPVC-LEPCITDIDHMRNNIOLKWKYEGK 475  
 QY 314 --HGRLYFESRRPYF--PVPICKEYSYCDNGFTTPSOSYMDLRCVTNGMEPEVPCL 368  
 DB 476 ILHGDLIDPYCKQKGYNLSPIPL-SELISACNRC-----DVRTPCIRKESKMCASPPVI 530  
 QY 369 RQCFHYVEYGE-SSYQWRRYIEGQSAVQCHSGSYSLPNGQDYTYCTENGSPPPKCV 425  
 DB 531 RN-----GDIVSSAARTYENGSSVEYRCFQDNHFL-QGSQNVYCVDGWMTTPPSC 579  
 RESULT 13  
 CFWL\_HUMAN STANDARD; PRT; 330 AA.  
 ID CFWL\_HUMAN  
 AC Q03591;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE COMPLEMENT FACTOR H-LIKE PROTEIN 1 PRECURSOR (H36).  
 GN HFL1 OR CFHL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RX MEDLINE: 91201892.  
 RA Estaller C., Koistinen V., Schwaible W., Dierich M.P., Weiss E.H.;  
 RT "Cloning of the 1.4-kb mRNA species of human complement factor H  
 RT reveals a novel member of the short consensus repeat family related  
 RT to the carboxy terminal of the classical 150-kDa molecule.",  
 RJ J. Immunol. 146:3190-3196(1991).  
 [2]  
 RP SEQUENCE OF 4-330 FROM N.A.  
 RX MEDLINE: 91268081.  
 RA Skerka C., Horstmann R.D., Zipfel P.F.;  
 RT "Molecular cloning of a human serum protein structurally related to  
 RT complement factor H.",  
 RJ J. Biol. Chem. 266:12015-12020(1991).  
 CC -1- FUNCTION: MIGHT BE INVOLVED IN COMPLEMENT REGULATION.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- TISSUE SPECIFICITY: LIVER.  
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
 CC -1- SIMILARITY: STRONG, TO FACTOR H.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M65292; AAA35946.1; -  
DR EMBL; M65293; AAA35947.1; -  
DR EMBL; X56209; CAA39666.1; -  
DR PIR; A40455; A40455.  
DR PIR; S14604; S14604.  
DR HSSP; P08603; 1HF1.  
DR MIM; 134371; -  
DR INTERPRO; IPR000436; -  
DR PFAM; PF00084; sush1; 5.  
DR Repeat; Glycoprotein, Sush1; Signal; Polymorphism.  
KW SIGNAL 1 18  
FT CHAIN 19 330 COMPLEMENT FACTOR H-LIKE PROTEIN 1.  
FT DOMAIN 22 328 4 X SUSHI (SCR) REPEATS.  
FT REPEAT 22 84 SUSHI 1.  
FT REPEAT 86 141 SUSHI 2.  
FT REPEAT 146 202 SUSHI 3.  
FT REPEAT 207 263 SUSHI 4.  
FT REPEAT 265 328 SUSHI 5.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 157 157 Y -> H.  
FT VARIANT 159 159 /FTID=VAR\_001980.  
FT VARIANT 159 159 V -> L.  
FT VARIANT 175 175 /FTID=VAR\_001981.  
FT VARIANT 175 175 O -> E.  
FT CONFLICT 71 71 T -> N (IN REF. 2).  
FT SEQUENCE 330 AA; 37661 MW; 8DCOD3F92A85E035 CRC64;

Query Match 12.8%; Score 315; DB 1; Length 330;  
Best Local Similarity 43.2%; Pred. No. 4e-17;  
Matches 51; Conservative 21; Mismatches 46; Indels 0; Gaps 0;

OY 307 CDPFQKNGHLYEESRPFYVPVPIGKEYSYCDNGFTFSQSYWYDLKRTVNGMEPEVP 366  
DB 23 CDEPKIHGILTYDEKPKPFSQVPTGEVYFSCENYFNVSFSPKSPWTRITCTEEGMSPTPK 82  
OY 367 CLNCGCFHHYVEGESSWQRRYIEGOSAKVOCHSGYSLPFGODTYCTENGMSPPRC 424  
DB 83 CLNLCFEPVENGHSESSGQTHLEGDVITICMGRYLNQNNNISCVGRGWSPTPKC 140

## RESULT 14

VCP\_VACCV STANDARD: PRT: 263 AA.  
ID P10998;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE COMPLEMENT CONTROL PROTEIN PRECURSOR (VCP) (SECRETORY PROTEIN 35)  
DE (PROTEIN C3) (28 KDA PROTEIN).  
GN C3L.  
OS Vaccinia virus (strain WR), and Vaccinia virus (strain Copenhagen).  
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
OC Orthopoxvirus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-37.  
RC STRAIN-WR;  
RC STRAIN-WR;  
RA MEDLINE; 88318974.  
RA Kotwal G.J., Moss B.;  
RT "Vaccinia virus encodes a secretory polypeptide structurally related  
RT to complement control proteins.";  
RL Nature 335:176-178(1988).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-WR;  
RC STRAIN-WR;  
RX MEDLINE; 89073756.  
RX Kotwal G.J., Moss B.;  
RT "Analysis of a large cluster of nonessential genes deleted from a  
RT vaccinia virus terminal transposition mutant.";  
RL Virology 167:524-537(1988).  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN-COPENHAGEN;  
RX MEDLINE; 91021027.  
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
RA Paoletti E.;  
RT "The complete DNA sequence of vaccinia virus.";  
RL Virology 179:247-266(1990).  
RN [4]  
RN COMPLETE GENOME.  
RC STRAIN-COPENHAGEN;  
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,  
RA Paoletti E.;  
RL Virology 179:517-563(1990).  
RN [5]  
RN FUNCTION.  
RX MEDLINE; 92115714.  
RA Isaacs S.N., Kotwal G.J., Moss B.;  
RT "Vaccinia virus complement-control protein prevents  
RT antibody-dependent complement-enhanced neutralization of infectivity  
RT and contributes to virulence.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:628-632(1992).  
RN [6]  
RN STRUCTURE BY NMR OF 146-263.  
RX MEDLINE; 97446168.  
RA Miles A.P., Shaw G., Bright J., Perczel A., Campbell I.D.,  
RA Barlow P.N.;  
RT "NMR studies of a viral protein that mimics the regulators of  
RT complement activation.";  
RL J. Mol. Biol. 272:253-265(1997).  
CC -1- FUNCTION: SERVES TO PROTECT THE VIRUS AGAINST COMPLEMENT ATTACK BY  
CC INHIBITING BOTH CLASSICAL AND ALTERNATIVE PATHWAYS OF COMPLEMENT  
CC ACTIVATION. BINDS C3B AND C4B.  
CC -1- SIMILARITY: BELONGS TO THE SUPERFAMILY OF THE REGULATORS OF  
CC COMPLEMENT ACTIVATION (RCA).  
CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X13166; CAA31564.1; -  
DR EMBL; M22812; AAA69605.1; -  
DR EMBL; M35027; AAA47997.1; -  
DR PIR; A31005; WMVZSP.  
DR PDB; 1VVC; 03-DEC-97.  
DR PDB; 1VVD; 03-DEC-97.  
DR PDB; 1VVE; 03-DEC-97.  
DR INTERPRO; IPR000436; -  
DR PFAM; PF00084; sush1; 4.  
KW Signal; Repeat; Sush1; 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 263 COMPLEMENT CONTROL PROTEIN.  
FT DOMAIN 20 262 4 X SUSHI (SCR) REPEATS.  
FT REPEAT 20 82 SUSHI 1.  
FT REPEAT 85 144 SUSHI 2.  
FT REPEAT 147 202 SUSHI 3.  
FT REPEAT 205 262 SUSHI 4.  
FT DISULFID 21 70 BY SIMILARITY.  
FT DISULFID 54 81 BY SIMILARITY.  
FT DISULFID 86 126 BY SIMILARITY.  
FT DISULFID 112 143 BY SIMILARITY.  
FT DISULFID 148 190 BY SIMILARITY.  
FT DISULFID 176 201 BY SIMILARITY.  
FT DISULFID 206 248 BY SIMILARITY.  
FT DISULFID 234 261 BY SIMILARITY.  
FT SEQUENCE 263 AA; 28629 MW; E4322CC9A6EF8997 CRC64;

Query Match 12.8%; Score 314.5; DB 1; Length 263;



•Tue Nov 21 16:57:31 2000

us-09-316-163-14.rsp

Page 17

```

0Y 50 NEMWP-----SNPERICKRRCCRRGOTPRGSPFLAVGSEFEENGAKVYTCDE 98
Db 187 PEFYPERBEELYKECGKVITPOHVLMN--CSHP-----LG-PFESNSOCTPSCAE 232
0Y 99 GYOLLGEIDRECDADG-WTNDIPICEYVKCLPTLENGRIYSGAEMPODEYFGQYVR 157
Db 233 GYELDGPGL-QCLASGIWTTNNPKCDVAQCSLEAPHHGM--ACMHPILAAFYDSCK 289
0Y 158 PECSNGFIEBQKHECHSENGSLMSNEKPOCVEIISCLPRVE--NGDGIILKPYKKENER 214
Db 290 FECOPGYRARSNTLHCHGSGOMSEPLPTCAIACEPPEIPIHOSMDCVSTGFGYNS 349
0Y 215 FOYKCOKEFYVKKERGDVAICTSG-WN-POPECEMTGLTYIPNGITPRRIKHRIIDEL 272
Db 350 CTFICABEFVLKAGMDALICADSGMTAPAPCEALQC--PEFP---VPSKAOVNCSDF 403
0Y 273 -----RYECKNGEYPATRSPVSKCTITG-WIPA-PROSLKPCDFPOFKHGRLEYEE 321
Db 404 GTLIYQSVCSFSCDEGSLIVGAS-VIRCLATGHHNGAPPCQAVSC-----APMHSPE 455
0Y 322 SRRRYFPYPIKE-----YSYCCNGTTSOSIWDYLKCTVNG-WEPEYPCLR--QC-- 371
Db 456 NGSMTCVOAPLGNSTYKSTOCFMCDEGFYLSGP---ERLDCSPSGHMTGTPTCETAIKCPG 512
0Y 372 IFHRYVEYG--ESSYWRRYIEGOSAKVOCHSGSYLPMGQOTYCTENG-WS-PPPC 424
Db 513 IF-APEGNIDLCSHVHGFQGSICHSCHNDEPEL-LGSMNEVECTYSGRMSAPPPYC 567

```

Search completed: November 21, 2000, 16:49:28  
Job time: 331 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:48:55 : Search time 70.54 Seconds  
(without alignments)  
566.546 Million cell updates/sec

Title: US-09-316-163-14

Perfect score: 2454  
Sequence: 1 EDCKGPPPRENSEILSGSMS.....DTYYCTENGWSPPCVRIK 428

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP:REMBL\_14:\*  
2: SP:archaea:\*  
3: SP:bacteria:\*  
4: SP:fungi:\*  
5: SP:human:\*  
6: SP:mammal:\*  
7: SP:mhc:\*  
8: SP:organelle:\*  
9: SP:phage:\*  
10: SP:plant:\*  
11: SP:protist:\*  
12: SP:virus:\*  
13: SP:vertebrate:\*  
14: SP:unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1713	69.8	449	4	Q14570
2	1242	50.6	669	6	Q28085
3	813	33.1	452	11	Q61407
4	811	33.0	303	11	Q61405
5	569.5	23.2	1053	13	Q91275
6	566	23.1	808	11	Q61408
7	478	19.5	330	4	Q9UJ16
8	466	19.0	539	11	Q63135
9	453	18.5	679	11	Q99254
10	447.5	18.2	522	6	Q28769
11	447	18.2	2014	6	Q29530
12	437	17.8	559	4	Q9UOV2
13	437	17.8	2039	4	Q16745
14	437	17.8	2489	4	Q16744
15	436	17.8	1911	6	Q29528
16	434	17.4	661	6	Q29531
17	428	17.3	533	11	Q08569
18	424	17.3	645	12	Q9WRU2
19	414.5	16.9	522	6	Q28797

20	409	16.7	497	11	Q63612	Q63612
21	396	16.1	579	11	Q60736	Q60736
22	385.5	15.7	1045	6	Q46545	Q46545
23	375	15.3	560	5	Q22328	Q22328
24	373.5	15.2	974	5	P91658	P91658
25	371.5	15.1	1124	5	Q9VVR4	Q9VVR4
26	371	15.1	1087	4	Q14212	Q14212
27	369.5	15.1	315	6	Q28770	Q28770
28	369.5	15.1	1032	4	Q13866	Q13866
29	346	14.1	657	4	Q14006	Q14006
30	341	13.9	417	11	Q35520	Q35520
31	332.5	13.5	331	4	Q92496	Q92496
32	329	13.4	1652	5	Q9VU09	Q9VU09
33	319	13.0	343	11	Q61406	Q61406
34	318	13.0	483	11	Q64735	Q64735
35	317	12.9	550	12	P88903	P88903
36	317	12.9	550	12	Q40912	Q40912
37	315	12.8	330	4	Q9UJ17	Q9UJ17
38	314.5	12.8	259	12	P87616	P87616
39	307	12.5	243	4	Q14310	Q14310
40	302.5	12.3	740	4	Q95508	Q95508
41	298	12.1	754	6	Q28290	Q28290
42	297.5	12.1	263	12	Q89859	Q89859
43	297.5	12.1	263	12	Q07033	Q07033
44	296.5	12.1	263	12	Q89076	Q89076
45	296.5	12.1	285	6	Q19121	Q19121

## ALIGNMENTS

RESULT 1  
ID Q14570 PRELIMINARY: PRT: 449 AA.  
AC Q14570; P78435;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE COMPLEMENT FACTOR H PRECURSOR.  
GN HF OR CFH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE: 88134059.  
RT Ripoché J., Day A.J., Harris T.J.R., Sim R.B.;  
RL "The complete amino acid sequence of human complement factor H.";  
RN Biochem. J. 249:593-602(1988).  
RN [2]  
RP SEQUENCE OF 226-449 FROM N.A.  
RX MEDLINE: 86169701.  
RA Kristensen T., Wetsel R.A., Jack B.F.;  
RT "Structural analysis of human complement protein H: homology with Cab binding protein, beta 2-glycoprotein I, and the Ba fragment of B2.";  
RN J. Immunol. 136:3407-3411(1986).  
RN [3]  
RP SEQUENCE OF 1-19 FROM N.A.  
RA Vik D.P., Williams S.A.;  
RT Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-9 FROM N.A.  
RA Dominguez O.;  
RT Theists (1993); Immunologia, Hospital Trias I Pujol, Spain.  
DR EMBL: X07523; CAA50403.1; -  
DR EMBL: M12383; AAS52013.1; -  
DR EMBL: U56979; AAB01987.1; -  
DR EMBL: Z29665; CAA82763.1; -  
DR HSSP: P10998; IYVC.  
DR INTERPRO: IPR000436; -  
DR PFM: PF00084; sush1; 7.  
KW SIGNAL.  
FT SIGNAL. 1 18 POTENTIAL.

FT	CHAIN	19	449	POTENTIAL.
SQ	SEQUENCE	449 AA;	5103 MW;	C2AD47F155343E3 CRC64;

Query Match	69.8%;	Score 1713;	DB 4;	Length 449;
Best Local Similarity	67.4%;	Pred. No. 3.3e-142;		
Matches 288;	Conservative 48;	Mismatches 91;	Indels 0;	Gaps 0

Qy	1	EDCGPPRENSSELLSSMSSEOLYSECTQATYCKRGPIYLTGTYVYCKNGEWPSPNSR	60
Db	19	EDCMELEPPRRMTTELITGSMWSDQTYPECTQATYCKRGYSRLGAVIMVCKRGGEVVALNPLR	78
Qy	61	ICRRPCGHPEDTFRGSFRLAVSGSEFEFGAKVYYTODEGYQLLIGELIDYRECDADGNTNDI	120
Db	79	KCOKRPGCHPDPTFRGTFTLLTGAVFEYGAVKAYATOCNEGYTOLLIGELINRECDTGDGNTDI	138
Qy	121	PICEVYVCLPTLELGNRIYSGAAEPDOEYFGQVYRFEBCNSGFTIEGOKEMHCSNGML	180
Db	139	PICEVYVCLPTLADENKRYVSSAAEPREYHFQOAVRFVONSQYKLEGDENHCSDGFW	198
Qy	181	SNKPQOCYEICLPPRYNGCITLYLKVPYKENERFOYCKQSGEYVYKERGDVATGSGMWP	240
Db	199	SKERPCKYEICSKSPDYINSGPISQKLIYENENEFQYCKMGMGEYBSRGDAVATESGMWP	258
Qy	241	QPSCEMTCPLTPYIPNGIYTPHRIKRIHIDEIRYECKNGEYPAITRSBVSCKTITGMIAP	300
Db	259	LPSEGEKSCDMPYIPNGYSPLRIKHTGTEIYOCNGEYPAITRNTACTACSTGMIAP	318
Qy	301	RCSLKPQCFQPFKRGRLYEESRBRPVPVPIGKEYSYCCNGEYTPSOSQWYDIACOTVNG	360
Db	319	RCTLKPQCFPIPKRGRLYEHENMRPYPVAVGKTYSTYCCDHEETFRSGSYWMDIHCTOG	378
Qy	361	WEPEYVCLROCIHFHYVEGESSYQWRRYIEGQSAKYVCHSGSYSLPNSQDTYYCTENGWSP	420
Db	379	WSPAPVCLRKCYFYLENGYQNQNGRRFVQGSIDVACHGYALPFAQTLTYTCMENGWSP	438
Qy	421	PPKCYRI 427	
Db	439	TPRCIRV 445	

RESULT	2		
Q28085			
ID	Q28085	PRELIMINARY;	PRT; 669 AA.
AC	Q28085;		
DT	01-NOV-1996	(TREMBLrel, 01, Created)	
DT	01-NOV-1996	(TREMBLrel, 01, last sequence update)	
DT	01-MAY-2000	(TREMBLrel, 13, last annotation update)	
DE	CCP MODULES 3-12, WITH PARTS OF CCP 2 AND 13 (FRAGMENT).		
OS	Bos taurus (Bovine).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;		
OC	Bovidae; Bovinae; Bos.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER;		
RX	MEDLINE; 96202005.		
RA	Soanes C.J., Day A.J., Sim R.B.;		
RT	"Prediction from sequence comparisons of residues of factor H involved		
RT	in the interaction with complement component C3b."		
RL	Biochem J. 315:523-531(1996).		
DR	EMBL; X98697; CAA67257.1; -.		
DR	HSSP; P10998; IVC.		
DR	INTERPRO; IPR000436; -.		
DR	PFAM; PF000084; sushi; 11.		
FT	NON_TER	1	
FT	NON_TER	669	669
SEQ	SEQUENCE	669 AA;	75683 MW; D0D9DB30EE747AC2 CRC64;

Query Match	50.6%;	Score 1242;	DB 6;	Length 669;
Best Local Similarity	58.4%;	Pred. No. 9.7e-101;		
Matches 206;	Conservative 57;	Mismatches 88;	Indels 2;	Gaps 1;

OY		76	GSEFLAVGSEEFKAKVVMCDGEYGLLGDIDYECCADQMTNDIPICEVVKCLPTYLE	1355
Db		3	GSPHIAESNOEYFAKVVYCDESTGYWGENFRFECTDNGTNDIPICEVVKCLPTIERE	62
OY		136	NGRTIVSGAAEEPDQEYGVGOVVRPECSNGFKIEGOKEMHSCENGLMSNEKPDCVEISCLPP	1955
Db		63	NGKIFSDALBEPDOGYTGVOVGFECNSGYMLDKPKQLHCSSAGVSAFEPRCKVEIFCKPP	1222
OY		196	RVENGDGIYLKPVYKENRFPYKKKGQGVYKKERDAVCCTGGSNRPQPSCEHMCIPLFYIP	2555
Db		123	VILNGOAVLPATKYKONERNVRYCRAAGFEYGORDDVTCTKSGMTAPATCLEITCDPRIP	1822
OY		256	NGIYTPHRIKRIDELREYCKNCFYPATRSVPKCTITGMIPARCSLKPCDFOPQRHG	3155
Db		183	NGVTRPELSKTRGODKLTYECKKGFPEPIKGTACTCRDGNVPRPCAMKPCSIPYIKHG	2422
OY		316	RLYYEESRRPFPPIGKEYSYSDNGFTTPSOSYWDYLRLCTVGMGEPEVPCLRQCIFHY	3755
Db		243	RLIY--SYRGTFPARVNQGFYSCHDHNVPRSQSWMHLACTAGWSPEBCJLRQCIFYN	3000
OY		376	VEYGESSYWRRIYEGOSAKYQCHSGSYSLPGNQDTYYCTENGSMSPPKCYARIK	428
Db		301	LENGHNQREBKLYGETRVACHCEGYSLDNQDMTMTCTESGMSPPRCIRVX	353
RESULT	3			
O61407		PREDIMINARY;	PRT;	452 AA.
O61407:				
AC		061407:		
PT		01-NOV-1996 (TREMBLrel_01, Created)		
PT		01-NOV-1996 (TREMBLrel_01, Last sequence update)		
DT		01-MAY-2000 (TREMBLrel_13, Last annotation update)		
DE		COMPLEMENT FACTOR H-RELATED PROTEIN.		
OS		Mus musculus (Mouse).		
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NC		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RA		Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,		
RA		Chaplin D.D.;		
RL		J. Biol. Chem. 0:0-0(0) .		
RN		[2]		
RP		SEQUENCE FROM N.A.		
RA		Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,		
RA		Chaplin D.D.;		
RL		Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.		
DR		EMBL: M29010; AAA37415.1; -.		
DR		HSSP: P08603; IHP1.		
DR		INTERPRO: IPR000436; -.		
DR		PFAM: PF00084; sushi.7.		
SQ		SEQUENCE 452 AA; 51602 MW; 2B697A4FFC6E13CA CRC64;		

```

RESULT      3
ID           061407
AC           061407;
DT           01-NOV-1996 (TREMBLrel. 01, Created)
DT           01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT           01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE           COMPLEMENT FACTOR H-RELATED PROTEIN.
OS           Mus musculus (Mouse).
OC           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
RN           [1]
RP           SEQUENCE FROM N.A.
RA           Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA           Chaplin D.D.;
RL           J. Biol. Chem. 0:0-0(0).
RN           [2]
RP           SEQUENCE FROM N.A.
RA           Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,
RA           Chaplin D.D.;
RL           submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
DR           EMBL: M29010; AAA37415.1; -.
DR           HSSP: P08603; 1HFI.
DR           INTERPRO: IPR000436; -.
DR           Pfam: PF00084; sushi_7
SQ           SEQUENCE 452 AA; 51602 MW; 26974AFCC613CA CRC64;

```

Query Match	33.18;	Score 813;	DB 11;	Length 452;
Best Local Similarity	71.08;	Pred. No. 2.3e-63;		
Matches 137;	Conservative 18;	Mismatches 38;	Indels 0;	Gaps 0

Qy	236	SCNMNQBPSCSEMTCLTLPYIPNGITYPHRIKHRIIDELRIECKNKQFVATSPSPVAKCITITG	285
Db	15	TAMLTSTANGAEETGCTSPPLITNGITYPHRIIHRSDDELRIECCNIGFTFYTGTSYVSKCIPTG	74
Qy	296	WIPARCSILPCDFEQFHGRILYEESSRRPYEPVPICKEYSYCGDNGFTTSPQSYWYL	355
Db	75	WIPVRCILTKCFEQQFQYGRILYEESSLRPNFVPSIGNKYSYRCDNGFSPSGSYWYL	134
Qy	356	CTVANGMEBPVPCLOCIITHYVEGESSYWRRIIEGOSAKYQCSAGSYLSPNGDTTYCTE	415
Db	135	CTAQGMEEPVPCVRCACVHYENGDSSAYWEKRIYQOGSLKYQCNYSYLONGODIMCTE	194
Qy	416	NGMSPPKCVIRK	428
Db	195	NGMSPPKCIIRK	207



RESULT 4  
ID 061405 PRELIMINARY; PRT: 303 AA.  
AC 061405;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT FACTOR H-RELATED PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90153969.  
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,  
Chaplin D.D.;  
RT "Identification and sequence analysis of four complement factor H-  
related transcripts in mouse liver."  
RL J. Biol. Chem. 265:3193-3201(1990).  
DR EMBL; M29007; AAA37413.1; -.  
DR HSSP; P10998; 1VVC.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 4.  
SQ SEQUENCE 303 AA; 34498 MW; 51C6E50906FAC24 CRC64;

Query Match 33.0%; Score 811; DB 11; Length 303;  
Best Local Similarity 70.8%; Pred. No. 2,2e-63;  
Matches 136; Conservative 19; Mismatches 37; Indels 0; Gaps 0;

QY 236 SGWNPQSCSEMTCLTPPIPGSTYPRHRIKRIIDETRECKNGFYPATRSVSKCTITG 295  
DB 15 TAMLSTANGEEKTSPYILNGLTYPHRIHRSDEIRKCYGYPYVGTSTVSKCTPTG 74  
QY 296 WIPAPRCSLKPCDPQPFHGRLYEESRRPYVPVIGKREYSYCDNGFTTSPQSWDYLR 355  
DB 75 WIPVRCTLKPCDPQPFHGRLYEESRRPYVPVIGKREYSYCDNGFTTSPQSWDYLR 134  
QY 356 CTVNGMEPEVPCIRQICIRHYVEGSSYWRRIYEGOSAKVOCHSGSLPNQODTYTCTE 415  
DB 135 CTAGQWMEPEVPCIRQICIRHYVEGSSYWRRIYEGOSAKVOCHSGSLPNQODTYTCTE 194  
QY 416 NGWSPPPKCVRI 427  
DB 195 NGWSPPPKCVRI 206

RESULT 5  
ID 091275 PRELIMINARY; PRT: 1053 AA.  
AC 091275;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT REGULATORY PLASMA PROTEIN.  
OS Paratubax nebulifer (barred sand bass).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii;  
OC Percormorpha; Perciformes; Percoidae; Serranidae; Paratubax.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-LIVER;  
RA MEDLINE; 94318039.  
RA Dahmen A., Kaidoh T., Zipfel P.F., Gagli I.;  
RT "Cloning and characterization of a cDNA representing a putative  
complement-regulatory plasma protein from barred sand bass (Paratubax  
nebulifer).";  
RL Biochem. J. 301:391-397(1994).  
DR EMBL; L21703; AAA92556.1; -.  
DR HSSP; P08603; 1HFN.  
DR INTERPRO; IPR000436; -.

DR PFAM; PF00084; sushi; 16.  
SQ SEQUENCE 1053 AA; 117597 MW; F27E32C3AD76D5D3 CRC64;

Query Match 23.2%; Score 569.5; DB 13; Length 1053;  
Best Local Similarity 26.7%; Pred. No. 1.4e-41;  
Matches 148; Conservative 60; Mismatches 169; Indels 177; Gaps 19;

QY 21 EQLYSEGTQATYKCRPGYTLCTIYKVCNKGEMV/NSPBRICKRRCGHPGTPGSEFL 80  
DB 47 EASYPGRGQVRVGCNNGYS--GFFKLVCVEGKMETRGAK--COPRSCGHPGDAQFADFL 102  
QY 81 AVGSEEEGAKVYVYCCDEYQTLGELIDYRECDADGNTNIPICEVYKCLPVTELENGRTV 140  
DB 103 AEGNDPFGSKVYVYCCQKGTOMVSRINRCVABEGMDGVVPCESQCC--PLIHVDNNVQY 161  
QY 141 SGAEPDQEVYFGVAVFPCNSGFKI--EGQKEMHSGENGLAWSNEKPCQVEISCLPVRVEN 199  
DB 162 IGGPE---EATFGNVVRFSCSKSRSEILDSPELYCDEBDMGSPVPKKAITCAIPPIEN 218  
QY 200 GGGIYUKPYKENERQYCKGQGFYKKEGDAVCTGSG---WNFOPSCSEMTCL----- 250  
DB 219 GNVPGAIRREYKENDVLYHECDRAFKHIDR--PSTCIKQIKABWSPPLCESIKCRLTIMD 277  
QY 251 -TPYIP-----NGIYT----- 260  
DB 278 GTRIEPAYNLFSPGETTLKVICARTSWISTPOETSVYTTCCQDNGEWSIRPTCOEVRCSNR 337  
QY 261 -PHRI-----KRIIDETRECKNGFYPATRSVSKCTITGNIAP----- 300  
DB 338 REHVDSMVRSMERYTLDDNTRFWCKRGYKRGVYMTATCGNMGWMPPLCEVKTCSKE 397  
QY 301 -----RCSLKPCDPQFHK 314  
DB 398 NIQDAVIVGTQKIYINAKAIYACGEGNRRGITTLTCEGNSGDRKCVKPCPLPKPD 457  
QY 315 GRLYEESRRPYVPVIGKREYSYCDNGFTTSPQSWDYLRCTVNGMEPEVPCIRQIC-- 372  
DB 458 NGFF---RGPY---TGRLVYTTCKDGYKRLTEGMMAKACVGYW-----PELTICISN 505  
QY 373 ---FHYVEGSSYWRRIYEGOSAKVOCHSGSLPNQODTYTCTENGW----- 418  
DB 506 TTGCKRPEIPNAEVI---RRYEVQTVQYICNGYS--TQANSFSENGWMLLYGLSPQD 560  
QY 419 -----SPPPK 423  
DB 561 ICTRADVCGPPPE 574

RESULT 6  
ID 061408 PRELIMINARY; PRT: 808 AA.  
AC 061408;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, last annotation update)  
DE COMPLEMENT FACTOR H-RELATED PROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90153969.  
RA Vik D.P., Munoz-Canoves P., Kozono H., Martin L.G., Tack B.F.,  
Chaplin D.D.;  
RT "Identification and sequence analysis of four complement factor H-  
related transcripts in mouse liver."  
RL J. Biol. Chem. 265:3193-3201(1990).  
DR EMBL; M29009; AAA37416.1; -.  
DR HSSP; P08603; 1HCC.  
DR INTERPRO; IPR000436; -.  
DR PFAM; PF00084; sushi; 13.  
SQ SEQUENCE 808 AA; 91654 MW; 6FD97D53CE74DF6D CRC64;



"The murine complement receptor gene family. IV. Alternative splicing of Cr2 gene transcripts predicts two distinct gene products that share homologous domains with both human CR2 and CRI.".

RT J. Immunol. 144:3581-3591(1990).

12

RA SEQUENCE OF 21-367 FROM N.A.

RA MEDLINE: 95105691.

RA Kim Y.U., Kinoshita T., Molina H., Hourcade D., Seta T., Wagner L.M., Holers V.M.;

RT "Mouse complement regulatory protein Cr2/p65 uses the specific mechanisms of both human decay-accelerating factor and membrane cofactor protein."

RT J. Exp. Med. 181:151-159(1995).

DR EMBL: U17126; AAA78271.1; JOINED.

DR EMBL: U17123; AAA78271.1; JOINED.

DR EMBL: U17124; AAA78271.1; JOINED.

DR EMBL: U17125; AAA78271.1; JOINED.

DR EMBL: U17126; AAA78271.1; JOINED.

DR EMBL: U17127; AAA78271.1; JOINED.

DR EMBL: M36470; AAA37449.1; -.

DR HSSP: P10998; IYVC.

DR MGD; MGI:88489; Cr2.

DR INTERPRO: IPR000436; -.

DR PFAM: PF00084; sush1; 10.

FT NON\_TER

FT SEQUENCE 679 AA; 74916 MW; 52FC0FDCE20CDC CRC64;

Query Match 18.5%; Score 453; DB 11; Length 679;

Best Local Similarity 28.6%; Pred. No. 1,3e-31;

Matches 126; Conservative 73; Mismatches 178; Indels 64; Gaps 27;

QY 24 YSEGTQATYKCRPGYRTLTGTYKCKNGEWPNSPRTCKRPGCHGDPFGSRLAVG 83

DB 42 FAIGTTWEYKCRPGFRKSFITILETSKW--SDAQCFKRRKPKMNPDEPLHGSVHINTG 99

QY 84 SFEEGAVVYTCDEGYOLLEIDYRECDAG--WTNDIPICEVVKCLPTELENGRIY 140

DB 100 --TEGSTYTYSCNNGYRLIGDSATCIVSDTYWMDMDPLCESTIPESPPAISNGDFY 157

QY 141 SGAAPDEQYVGFVFECSNG-----FKIEGQEMHCSEN-----GLMSNEKPOCV-E 189

DB 158 SSSRD---SFGYGVVYTYCHGKMKREKFLPDVGEKSTYCTSKDQVIMWSPRQICPR 214

QY 190 ISCLPPEVNG--DGYLKPYKENERFYKCKGQGYKKEGDAVC--TSGMNPQ--PSC 244

DB 215 VKCPMPPEIENGIVESG--FKHSFPLNDYVIRCKSGFTMKSRIMWCPDNKMSPPLEPTC 272

QY 245 EEMPTCLTPY-IPNGITYPTHRIKRIHIDELRECKNGFYPATRSPVSKCTITG-WI-PAPR 301

DB 273 -FMGCLPPONIIHGDIYNNKDEFFSYGOKVSYTCNPG-YTLIGTNLVECTSLGTSNIVPT 330

QY 302 CSLKPCD-FP-QFKHGRLYEESRRPYFPVPIGKEYSYCDNGFT--TPSQSYWDLRC 356

DB 331 CEVKSQDAIPIHNLHGRVFLPNN-----LQGAEVSPVCDLGFLKGPSS-----QC 378

QY 357 TVNG--WEPEVPCLRQICF--YVEYGESYQWRKRIEQAQKQVCHSISYLPNQD 409

DB 379 IPEGETVIMNNKFPVCEQISCDPPEVKNARKPYSLPIVGTALRYCSPSYRL-IGEK 437

QY 410 TYVC-TEN---GM-SPPK 424

DB 438 AIFCISENVHATWDKAPPIC 458

RESULT 10

028769 PRELIMINARY: PRT: 522 AA.

AC 028769;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)

DE COMPLEMENT RECEPTOR (FRAGMENT).

OS Papio hamadryas (Hamadryas baboon).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

OC Cercopithecoidea; Papio.

11

RA SEQUENCE FROM N.A.

RA TISSUE-BONE MARROW;

RA Birmingham D.J., Logar C.M., Shen X.P., Chen W.;

RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL: L77977; AAA99004.1; -.

DR HSSP: P08603; IHR.

DR INTERPRO: IPR000436; -.

DR PFAM: PF00084; sush1; 7.

FT NON\_TER

FT SEQUENCE 522 AA; 56626 MW; 312FCBE03ADF19DC CRC64;

Query Match 18.2%; Score 447.5; DB 6; Length 522;

Best Local Similarity 29.7%; Pred. No. 2,9e-31;

Matches 134; Conservative 66; Mismatches 168; Indels 83; Gaps 31;

QY 27 GTQATYKCRPGYRTLTGTYKCKNGEWPNSPRTCKRPGCHGDPFGSRLAVGSEF 86

DB 71 GYLYKTECLPGYHKGKPFSLICLNKSWTSADK--CTRKSCKRMPKDPVNG--MVAHIXDI 126

QY 87 EFGAKVYVTCDEGYOLLEIDYRECDAG--WTNDIPICEVVKCLPTELENGRIYS 141

DB 127 QFQSQIYSCNKGIRLIGS--SSATCIISGNTVIMDETPICETIPCGLEPT-IANGDFIS 184

QY 142 GAAPDEQYVGFVFECSNG-----FKIEGQEMHCSEN-----GLMSNEKPOCV-VE 189

DB 185 ---TSREYFPYGSVYTRCNLISGRRKFLFELVGEPSYCTSKDDQVIGMSPAPQCIIP 240

QY 190 ISCLPPEVNGDGYL-KPYKENERFYKCKGQGYKKEGDAVCTG--SGMNPQ--PSC 246

DB 241 NKCPMPPEVNGIVYSVRSLSFSLNEVEFRQCPFYWKGRHQQCALNKWPELPSCSR 300

QY 247 MTCCLTPYIPNGITYPTHRIKRIHIDELRECKNGFYPATRSPVSKCTITG-WI-APRCSL 304

DB 301 VCPPEILGHEHTPSQODPSPGQEVYSCPEG-YDLRGAASLIHCTIPQGDMPNPAITCY 359

QY 305 KPCD--FPQFKHGRLYEESRRPYFP--VPYIGKEYSYCDNGFTTPSQSYWDLRCTVNG 360

DB 360 KSCDFFLGQLPHGRV-----LFLNLQGAKVSVCDEGRFLKGRF--ASHCVLAG 408

QY 361 ---WEPEVPCLRQICF-----HY-VEYGESYQWRKRIEQAQKQVCHSISYLPNQD 402

DB 409 MKALMNSVVPCEQ-IFCPNPPIALNGRHIGAPLDIPY-----GKEYSYICDHPDR 460

QY 403 SLP---NGDHYCTE---NG-W-SPPK 424

DB 461 GMTVNLIGESTIRCTSDPQNGWSSPAPRC 491

RESULT 11

029530 PRELIMINARY: PRT: 2014 AA.

AC 029530;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)

DE COMPLEMENT RECEPTOR 1 (FRAGMENT).

GN CRI.

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

11

RA SEQUENCE FROM N.A.

RA Birmingham D.J., Shen X.P., Hourcade D., Nickells M.W., Atkinson J.P.;

RT "Primary sequence of an alternatively spliced form of CRI. Candidate for the 75,000 M(r) complement receptor expressed on chimpanzee erythrocytes."

RESULT	12	
09UUV2		
ID	09UUV2,	PRELIMINARY;
AC	09UUV2,	PRT; 559 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)	
DE	HUMAN C3B1 mRNA FOR C3B/44B RECEPTOR SECRETED FORM PRECURSOR (FRAGMENT)	
OS	Homo sapiens (Human)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE: 89010527.	
RA	Houcraie D., Wiesner D.R., Atkinson J.P., Holers V.M.;	
RT	"Identification of an alternative polyadenylation site in the human	
RT	C3b/C4b receptor (complement receptor type 1) transcriptional unit and	
RT	prediction of a secreted form of complement receptor type 1.";	
RL	J. Exp. Med. 168: 1255-1270(1988).	
DR	EMBL: X14362: CAA325A1.1; -.	
DR	INTERPRO: IPR000436; -.	
DR	INTERPRO: IPR000834; -.	

ID	016745	PRELIMINARY;	PRT;	2039	AA.
AC	016745;				
DT	01-NOV-1996 (TREMBLrel. 01, Created)				
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)				
DE	COMPLEMENT RECEPTOR 1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	MEDLINE: 94065175.				
RT	Vik D.P., Wong W.W.;				
RT	"Structure of the gene for the F allele of complement receptor type 1				
RT	and sequence of the coding region unique to the S allele.";				
RT	J. Immunol. 151:6214-6224(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RP	Vik D.P., Wong W.W.;				
RA	Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.				
RL	EMBL, L17418; AAB60694.1; -				
DR	EMBL, L17390; AAB60694.1; JOINED.				
DR	EMBL, L17391; AAB60694.1; JOINED.				
DR	EMBL, L17392; AAB60694.1; JOINED.				
DR	EMBL, L17393; AAB60694.1; JOINED.				

```

Db      432 HGKPLEVFPFKAVNYNCDDPHDGTGTFDLIGESTINCTSDPGNGVWSSPARC 487

RESULT  14
ID      016744      PRELIMINARY;      PRT; 2489 AA.
AC      016744;
DT      01-NOV-1996 (TREMblrel. 01, Created)
DT      01-NOV-1996 (TREMblrel. 01, last sequence update)
DE      01-MAY-2000 (TREMblrel. 13, last annotation update)
DE      COMPLEMENT RECEPTOR 1.
OS      Homo sapiens (Human) .
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
[1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 94065175.
RA      Vik D.P., Wong W.W.;
RT      "Structure of the gene for the F allele of complement receptor type 1
RT      and sequence of the coding region unique to the S allele.";
RN      J. Immunol. 151:6214-6224(1993).
[2]
RP      SEQUENCE FROM N.A.
RA      Vik D.P., Wong W.W.;
RL      Submitted (JUN-1993) to the EMBL/GenBank/DDBJ databases.
RL      EMBL; L17418; AAB60695.1; -.
DR      EMBL; L17390; AAB60695.1; JOINED.
DR      EMBL; L17391; AAB60695.1; JOINED.
DR      EMBL; L17392; AAB60695.1; JOINED.
DR      EMBL; L17393; AAB60695.1; JOINED.
DR      EMBL; L17394; AAB60695.1; JOINED.
DR      EMBL; L17395; AAB60695.1; JOINED.
DR      EMBL; L17396; AAB60695.1; JOINED.
DR      EMBL; L17397; AAB60695.1; JOINED.
DR      EMBL; L17398; AAB60695.1; JOINED.
DR      EMBL; L17399; AAB60695.1; JOINED.
DR      EMBL; L17400; AAB60695.1; JOINED.

```

```

DR EMBL: L17402; AAB60695.1; JOINED.
DR EMBL: L17403; AAB60695.1; JOINED.
DR EMBL: L17405; AAB60695.1; JOINED.
DR EMBL: L17406; AAB60695.1; JOINED.
DR EMBL: L17407; AAB60695.1; JOINED.
DR EMBL: L17408; AAB60695.1; JOINED.
DR EMBL: L17409; AAB60695.1; JOINED.
DR EMBL: L17410; AAB60695.1; JOINED.
DR EMBL: L17411; AAB60695.1; JOINED.
DR EMBL: L17413; AAB60695.1; JOINED.
DR EMBL: L17413; AAB60695.1; JOINED.
DR EMBL: L17414; AAB60695.1; JOINED.
DR EMBL: L17415; AAB60695.1; JOINED.
DR EMBL: L17415; AAB60695.1; JOINED.
DR EMBL: L17417; AAB60695.1; JOINED.
DR EMBL: L17419; AAB60695.1; JOINED.
DR EMBL: L17420; AAB60695.1; JOINED.
DR EMBL: L17421; AAB60695.1; JOINED.
DR EMBL: L17422; AAB60695.1; JOINED.
DR EMBL: L17423; AAB60695.1; JOINED.
DR EMBL: L17425; AAB60695.1; JOINED.
DR EMBL: L17426; AAB60695.1; JOINED.
DR EMBL: L17427; AAB60695.1; JOINED.
DR EMBL: L17428; AAB60695.1; JOINED.
DR EMBL: L17429; AAB60695.1; JOINED.
DR EMBL: L17430; AAB60695.1; JOINED.
DR HSSP: P08603; IHP1.
DR INTERPRO: IPR000436; -.
DR INTERPRO: IPR000834; -.
DR INTERPRO: IPR001424; -.
DR PFAM: PF00084; sushi; 37.
DR PROSITE: PS00087; SOD_CU_ZN.1; UNKNOWN_1.
DR PROSITE: PS00113; CARBOXYPEPT_ZN.2; UNKNOWN_3.
DR SQUENCE 2489 AA; 27846 MW; CEELILBS3FZB4EAF6 CRC64A

```

Query Match	17.8%;	Score 437;	DB 4;	Length 2489;
Best Local Similarity	27.7%;	Pred. No. 1.6e-29;		
Matches 132;	Conservative 63;	Mismatches 149;	Indels 132;	Gaps 31

Qy	27	GTQATVYKCRPEYRFLGIVKCKNGEWWFSPSRICKRRCQGHGDDPFSFLAAVSEF	86
Db	66	GTLYNDECRPEYSGRPPSIICLKNSVMTGAKDR--CARKSCBNPDPEVNMCMVHYKG--I	121
Qy	87	EFKAKVYITDECGEQLLAGEIDRECCDADG---WTNDIPICEVVKC-LPTYLENGIAYS	141
Db	122	QFGSQIKRSCFKCYRLLGS--SSATCIISGDVIMDNETPITCDIRIPGELPPT-ITNGEIS	179
Qy	142	GAAPDEDDYYGYGVVRFRECNCG-----FKIEGQKEMHCSEN-----GLAMENKPOC-VEI	190
Db	180	TNRE---NFHIGSVYTYRCNPGSGGRKRVFELVEPSPSYCTCSNDQVAYSINGAPAOCTIPN	236
Qy	191	SCLEPPRVENG---DGILYKLPVYKENERQYKCKQGFVYKREGDVCTG--SGWNPQ--PSC	244
Db	237	KCTPPNVENGILVSDN---RSLFSLINEVEFERCQPGFVYMKGRPRVYKCOALINKWPELPSC	293
Qy	245	EEAMCTLPYIINGIYTHRIKHKRD-----DEIRECKNGFYPARRSVNSCTITG--HIP-	298
Db	294	SRV--CQRP--PDVLYHAERTQDKDNKESPGOGVAFVYSCENG-YDLGASMSRCTPOGDMSPA	349
Qy	299	APRCSILKPCD--PQFKHGRILYEESRKRPVY--PIGKEYSYXCDAAGFTT--PSQSYWD	352
Db	350	APLCEVYKCDDEPMQOLNGRV-----LEFVNLQIGAKVDFVCDGPFQJLKSSAY--	399
Qy	353	YLKCTVNG---WEPEVPLKQCLIFHYVEGESSYWMORRYIEGOSAKVYQCHSGYSLPN--	406
Db	400	---CVLAGMESILMNSVYVCEQ-----IFCSPSPVIVNGR	431
Qy	407	-----GDDTYCYE-----NG-W--SPPKC	424
Db	432	HTGKPLVEYFPFGKAVNYTCDPHDPRGISFDLIGESTIRKCSIDPDGNGWSSAPARC	487

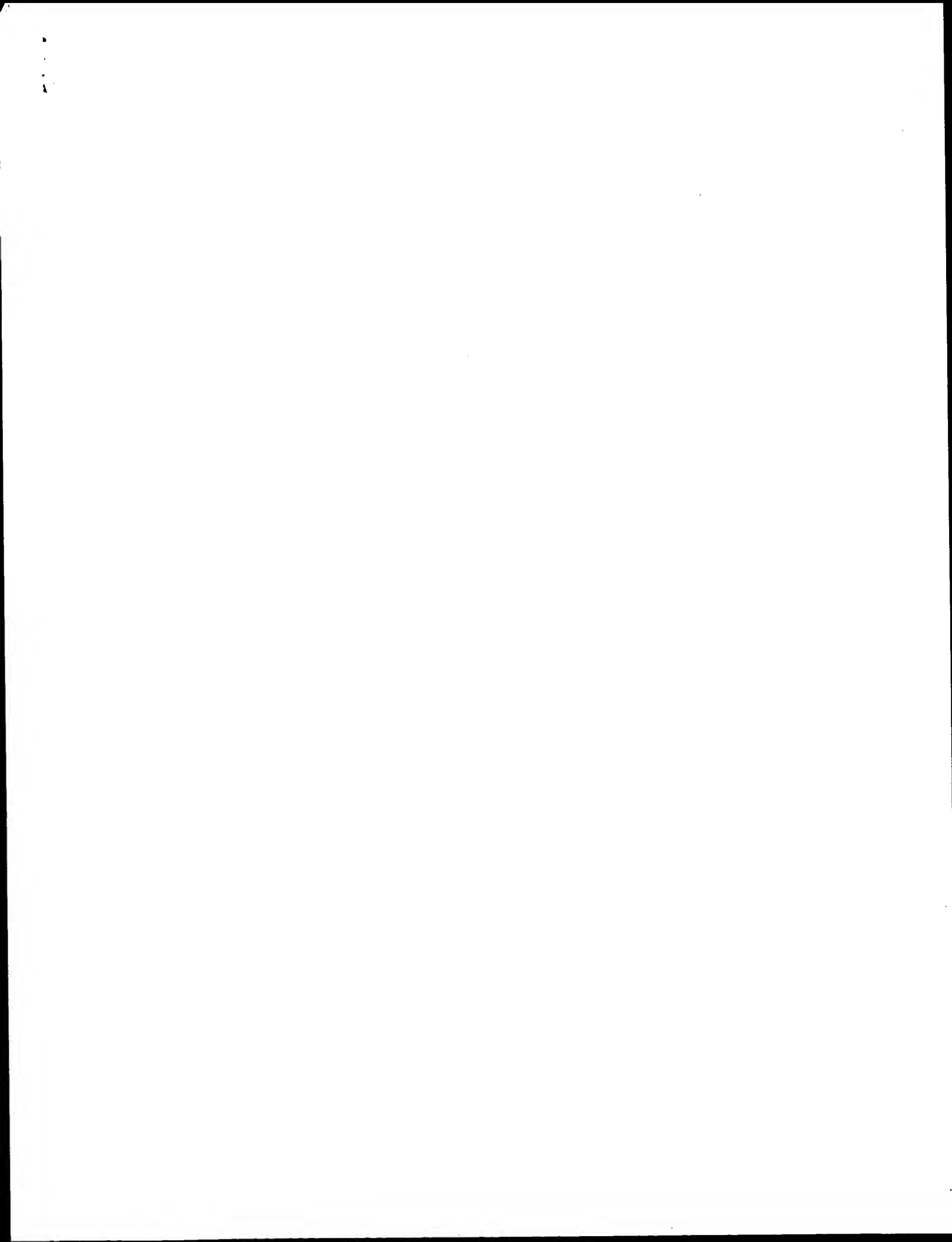
RESULT	ID	PRELIMINARY;	PRT;	1911 AA.
029528	Q29528			
AC	Q29528;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	COMPLEMENT RECEPTOR 1 (FRAGMENT).			
GN	CR1.			
OS	Papio hamadryas (Hamadryas baboon).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;			
OC	Cercopithecinae; Papio.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Clemenza L., Subramanian B.V., Nickells M.W., Hourcade D.E.,			
RA	Atkinson J.P., (MAR-1995) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; L39791; AAA6210.1; -.			
DR	HSSP; P08603; IHCC.			
DR	INTERPRO: IPR000436; -.			
DR	INTERPRO: IPR000834; -.			
DR	INTERPRO: IPR001424; -.			
DR	PFAM; PFO0084; sushi. 29.			
DR	PROSITE; PS00087; SOD_CU_ZN.1; UNKNOWN.1.			
DR	PROSITE; PS00133; CARBOXYPEPT_ZN.2; UNKNOWN.4.			
FT	NON_TER 1911 1911			
SQ	SEQUENCE 1911 AA; 210173 MW; 53564DD0EAF5A521D CRC64;			

Query Match	17.8%;	Score 436;	DB 6;	Length 1911;
Best Local Similarity	28.6%;	Pred. No. 1.4e-29;		
Matches 130;	Conservative 68;	Mismatches 168;	Indels 88;	Gaps 28

[illegible]

Search completed: November 21, 2000, 16:49:00  
Job time: 338 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:37 ; Search time 47.7 Seconds

(Without alignments)  
437.703 Million cell updates/sec

Title: US-09-316-163-11

Perfect score: 1876

Sequence: 1 EDNCNLPKRRMTTEILTGSMS.....PDIKHGCLYHMKRRPYPPV 329

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1876	100.0	449	1 NBHHS	complement factor
2	1876	100.0	1231	1 NBHHS	complement factor
3	1261	67.2	1234	1 NBHHS	complement factor
4	903	48.1	669	2 S65551	factor H - bovine
5	486	25.9	1053	2 S46199	probable complemen
6	356	19.0	597	1 S53711	C4BP alpha chain p
7	353	18.8	597	1 NBHHS	C4b-binding protei
8	353	18.8	676	2 A45300	complement C3b rec
9	350.5	18.7	2489	2 I73012	complement C3b/C4b
10	348	18.6	2014	2 I36936	complement recepto
11	343.5	18.3	482	2 A34924	complement C3b/C4b
12	337.5	18.0	560	2 T16833	complement C3b/C4b
13	335	17.9	360	2 T42821	hypothetical prote
14	323	17.2	497	2 JC2054	complement control
15	320.5	17.1	610	1 I46001	complement regulat
16	317	16.9	452	2 A35068	C4b-binding protei
17	317	16.9	558	2 S57953	complement factor
18	311.5	16.6	469	1 NBHHS	C4BP protein alpha
19	307	16.4	433	2 A30550	C4b-binding protei
20	306	16.3	303	2 H35068	complement C3b/C4b
21	304	16.2	440	2 A43519	apolipoprotein H-r
22	303.5	16.2	1025	2 A43526	complement recepto
23	303	16.2	661	1 KFHU13	complement C3d/Eps
24	301.5	16.1	263	1 MMVZSP	coagulation factor
25	294	15.7	668	2 A46013	apolipoprotein H h
26	288	15.4	1091	1 PLO009	coagulation factor
27	286	15.2	830	2 A30359	complement C3d/Eps
28	285	15.2	579	2 A56740	p-selectin precurs
29	284.5	15.2	263	1 C36838	sperm-egg recognit
					complement control

30	284.5	15.2	263	2 T28450	hypothetical prote
31	283.5	15.1	263	2 B72152	B18L protein - var
32	279.5	14.9	345	1 NBHHS	apolipoprotein H p
33	279	14.9	610	2 A35046	E-selectin precurs
34	276.5	14.7	612	2 S23174	endothelial leukoc
35	276	14.7	618	2 B42755	E-selectin precurs
36	274.5	14.6	349	2 G02913	sperm CD46 - human
37	274.5	14.6	369	2 I57998	membrane cofactor
38	274.5	14.6	768	2 A42755	p-selectin precurs
39	274	14.6	377	2 I54479	membrane cofactor
40	274	14.6	384	2 S01896	membrane cofactor
41	274	14.6	768	2 I53821	p-selectin rat
42	270.5	14.4	362	2 JC5194	membrane cofactor
43	270.5	14.4	369	2 JC5198	membrane cofactor
44	269.5	14.4	360	1 WMBE2E	membrane-bound com
45	267.5	14.3	302	1 WMBE1E	secretory compleme

#### ALIGNMENTS

RESULT 1  
NBHHS  
complement factor H precursor, short splice form - human  
N: Alternate names: complement factor H-related protein; complement protein H  
C: Species: Homo sapiens (man)  
C: Date: 31-Dec-1993 #sequence-revision 23-Feb-1996 #text-change 22-Jun-1999  
C: Accession: S03013; B60238; A27877; A61103; A26505; S10479  
R: Riipoche, J.; Day, A.J.; Harris, T.J.R.; Slim, R.B.  
Biochem. J. 249, 593-602, 1988  
A: Title: The complete amino acid sequence of human complement factor H.  
A: Reference number: S00254; MUID:88134059  
A: Accession: S03013  
A: Molecule type: mRNA  
A: Residues: 1-449 <RIP>  
A: Cross-references: EMBL:X07523; EMBL:Y00716; NID:932492; PIDN:CA30403.1; PID:G75807  
A: Note: part of this sequence, including the amino end of the mature protein was conf  
A: Note: 402-Tyr was also found  
R: Estaller, C.; Schwaebler, W.; Dierich, M.; Weiss, E.H.  
Eur. J. Immunol. 21, 799-802, 1991  
A: Title: Human complement factor H: two factor H proteins are derived from alternativ  
A: Reference number: A60238; MUID:91184292  
A: Accession: B60238  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 1-33;434-449 <EST>  
A: Note: only portions of this 1.8 kilobase mRNA were sequenced  
R: Schulz, T.F.; Schwaebler, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.  
Eur. J. Immunol. 16, 1351-1355, 1986  
A: Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequenc  
A: Reference number: A27877; MUID:87054207  
A: Accession: A27877  
A: Molecule type: mRNA  
A: Residues: 1L',55-401,'Y',403-449 <SCH>  
A: Note: an additional nucleotide present within the codon for Glu-310 was thought to  
R: Schwaebler, W.; Zwirner, J.; Schult, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.  
Eur. J. Immunol. 17, 1485-1489, 1987  
A: Title: Human complement factor H: expression of an additional truncated gene produc  
A: Reference number: A61103; MUID:88055295  
A: Accession: A61103  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 27-76 <SC2>  
A: Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that  
R: Slim, R.B.; Discipio, R.G.  
Biochem. J. 205, 285-293, 1982  
A: Title: Purification and structural studies on the complement-system control protein  
A: Reference number: A26505; MUID:83048213  
A: Accession: A26505  
A: Molecule type: protein  
A: Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>  
R: Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
Biochemistry 31, 3626-3634, 1992

A:Title: Solution structure of the fifth repeat of factor H: A second example of the com  
 A:Reference number: A44551; MUID:92232649  
 A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Kristensen, T.; Wetzel, R.A.; Tack, B.F.  
 J. Immunol. 136, 3407-3411, 1986  
 A:Title: Structural analysis of human complement protein H: homology with C4b binding pr  
 A:Reference number: S10479; MUID:86169701  
 A:Accession: S10479  
 A:Molecule type: mRNA  
 A:Residues: 226-401, 'Y', 403-449 <KRI>  
 A:Cross-references: GB:M12383; NID:q180472; PIDN:AA52013.1; PID:q180473  
 C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc  
 C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.  
 C:Genetics: <HF1>  
 A:Gene: GDB:HF1; HF  
 A:Cross-references: GDB:120041; OMIM:134370  
 A:Map position: 1q32-1q32  
 C:Genetics: <HF2>  
 A:Gene: GDB:HF2; HF  
 A:Cross-references: GDB:129095  
 A:Map position: 1q32-1q32  
 A:Note: The correspondence between the two loci and the sequences indicated is unclear;  
 C:Function:  
 A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa  
 be alternative complement pathway  
 A:Pathway: complement alternate pathway  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <Sig>  
 F:19-449/Product: complement factor H, short splice form #status experimental <MAT>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357  
 F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1876; DB 1; Length 449;  
 Best local Similarity 100.0%; Pred. No. 3, 1e-127;  
 Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCNELPRRNTETLLTGSWSQDTYPEGTOAIYKCRPGYRSLGNVYVCRNGEVALNPLR 60  
 |||||  
 DB 19 EDCNELPRRNTETLLTGSWSQDTYPEGTOAIYKCRPGYRSLGNVYVCRNGEVALNPLR 78  
 |||||  
 QY 61 KCKRPGHGDPRPEFTLTGCVNFEYKAVYTCNMGVOLLGEINVRCDTDGWNDI 120  
 |||||  
 DB 79 KCKRPGHGDPRPEFTLTGCVNFEYKAVYTCNMGVOLLGEINVRCDTDGWNDI 138  
 |||||  
 QY 121 PICEVYKCLPVTAPENKIIYSAMEPDREYHFGQAVFVCSNGYKIEGDEMHCSDDGFW 180  
 |||||  
 DB 139 PICEVYKCLPVTAPENKIIYSAMEPDREYHFGQAVFVCSNGYKIEGDEMHCSDDGFW 198  
 |||||  
 QY 181 SKRKPCKVEISCKSPDYINSPIISQKIIYKENERFOYKCMNGYTSRGAADVCTESSGWRP 240  
 |||||  
 DB 199 SKRKPCKVEISCKSPDYINSPIISQKIIYKENERFOYKCMNGYTSRGAADVCTESSGWRP 258  
 |||||  
 QY 241 LPSCSEKSCNPNYPNDYSPRLRIKHTGDEITYQCRNGFYPATRGMTAKTSGWIPAP 300  
 |||||  
 DB 259 LPSCSEKSCNPNYPNDYSPRLRIKHTGDEITYQCRNGFYPATRGMTAKTSGWIPAP 318  
 |||||  
 QY 301 RCTLPCDYPDIKRGHGLYHENMRPPPV 329  
 |||||  
 DB 319 RCTLPCDYPDIKRGHGLYHENMRPPPV 347  
 |||||

RESULT 2  
 NBIUH

complement factor H precursor, long splice form - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence-revision 31-Dec-1993 #text-change 22-Jun-1999  
 C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298  
 R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
 Biochem. J. 249, 593-602, 1988  
 A:Title: The complete amino acid sequence of human complement factor H.  
 A:Reference number: S00254; MUID:88134059  
 A:Accession: S00254  
 A:Molecule type: mRNA  
 A:Residues: 1-1231 <RIP>  
 A:Cross-references: EMBL:Y00716; NID:q31964; PIDN:CAA68704.1; PID:q31965  
 A:Note: 402-Tyr was also found  
 A:Note: parts of this sequence, including the amino and carboxyl ends of the mature p  
 R:Estaller, C.; Schwaebler, W.; Dietrich, M.; Weiss, E.H.  
 Eur. J. Immunol. 21, 799-802, 1991  
 A:Title: Human complement factor H: two factor H proteins are derived from alternat  
 A:Reference number: A60238; MUID:91184292  
 A:Accession: A60238  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-56,1177-1231 <EST>  
 A:Note: only portions of this 4.3 kilobase mRNA were sequenced  
 R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.  
 Biosci. Rep. 7, 201-207, 1987  
 A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl  
 A:Reference number: A54726; MUID:88025472  
 A:Accession: A54726  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 'DFRN', 579-1231 <DAY>  
 A:Cross-references: GB:M17517; NID:q180497; PIDN:AA52016.1; PID:q180498  
 A:Note: parts of this sequence were determined by protein sequencing  
 R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.  
 Biosci. Rep. 6, 65-72, 1986  
 A:Title: Partial characterization of human complement factor H by protein and cDNA se  
 A:Reference number: A61565; MUID:86188123  
 A:Accession: A61565  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 'METGRNHLMAK', 1050-1057, 'T', 1059-1102 <RI2>  
 R:Sim, R.B.; Discipio, R.G.  
 Biochem. J. 205, 285-293, 1982  
 A:Title: Purification and structural studies on the complement-system control protein  
 A:Reference number: A26505; MUID:83048213  
 A:Accession: A26505  
 A:Molecule type: protein  
 A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>  
 R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
 Biochemistry 31, 3626-3634, 1992  
 A:Title: Solution structure of the fifth repeat of factor H: A second example of the  
 A:Reference number: A44551; MUID:92232649  
 A:Contents: annotation; NMR structure determination, residues 264-292  
 R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.  
 J. Mol. Biol. 219, 717-725, 1991  
 A:Title: Three-dimensional structure of a complement control protein module in soluti  
 A:Reference number: A49224; MUID:91278097  
 A:Contents: annotation; NMR structure determination, residues 927-985  
 R:Estaller, C.; Koistinen, V.; Schwaebler, W.; Dietrich, M.P.; Weiss, E.H.  
 J. Immunol. 146, 3190-3196, 1991  
 A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no  
 A:Reference number: I56100; MUID:91201892  
 A:Accession: I72654  
 A:Status: translated from GB/EMBL/DBD  
 A:Molecule type: mRNA  
 A:Residues: 1047-1231 <RES>  
 A:Cross-references: GB:M65294; NID:q183766; PIDN:AA35948.1; PID:q183767  
 R:Carion, J.A.; Bates, R.C.; Smith, A.I.; Teto, T.; Arellano, A.; Gordon, D.L.; Burn  
 Biochim. Biophys. Acta 1289, 305-311, 1996  
 A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.  
 A:Reference number: S66298; MUID:96205365  
 A:Accession: S66298

```
A:Status: preliminary
A:Molecule type: protein
A:Residues: 411-419:574-578,580-582 <CAR>
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Hc
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.
C:Genetics: <HF1>
A:Gene: GDB:HF1; HF
A:Cross-references: GDB:120041; OMIM:134370
A:Map position: 1q32-1q32
C:Genetics: <HF2>
A:Gene: GDB:HF2; HF
A:Cross-references: GDB:129095
A:Map position: 1q32-1q32
A:Note: the correspondence between the two loci and the sequences indicated is unclear.
C:Function:
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa
he alternative complement pathway
A:Pathway: complement alternate pathway
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <StG>
F:19-449/Product: complement factor H #status experimental <MPF>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-267/Domain: complement factor H repeat homology <FH04>
F:246-328/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-623/Domain: complement factor H repeat homology <FH10>
F:630-684/Domain: complement factor H repeat homology <FH11>
F:691-744/Domain: complement factor H repeat homology <FH12>
F:753-803/Domain: complement factor H repeat homology <FH13>
F:811-864/Domain: complement factor H repeat homology <FH14>
F:870-926/Domain: complement factor H repeat homology <FH15>
F:931-984/Domain: complement factor H repeat homology <FH16>
F:989-1043/Domain: complement factor H repeat homology <FH17>
F:1048-1102/Domain: complement factor H repeat homology <FH18>
F:1109-1163/Domain: complement factor H repeat homology <FH19>
F:1167-1228/Domain: complement factor H repeat homology <FH20>
F:21-66:52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357
F:1-803,811-853,839-864,870-915,901-926,931-973,959-984,989-1032,1018-1043,1048-1091,1077-
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent
F:529,802,822,882,911/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:718,1029,1095/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match          100.0%: Score 1876; DB 1; Length 1231;
Best Local Similarity 100.0%: Pred. No. 9.2e-127;
Matches 329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EDNCNELPPRRNTEITLGSNSDQTYEGTQAIYKCRPGYSLGNVIMVCKKGVALNPIR 60
DB 19 EDNCNELPPRRNTEITLGSMSDQTYEGTQAIYKCRPGYSLGNVIMVCKKGVALNPIR 78
OY 61 KCQKRPCCGHPGCTPTLTGTGNVFEYGVAVYTCNBSGYLLGELINRECDTGTNDI 120
DB 79 KCQKRPCCGHPGCTPTLTGTGNVFEYGVAVYTCNBSGYLLGELINRECDTGTNDI 138
OY 121 PICEVVKCLPVPAPNGKIVSSAMEPDEYHFGQAVRFVPCNSGYITEGEENHCSDDGW 180
DB 139 PICEVVKCLPVPAPNGKIVSSAMEPDEYHFGQAVRFVPCNSGYITEGEENHCSDDGW 198
OY 181 SKERKVCVEISCKSPDVINGSPIQKIYKENERPOYKCNMGYSEKSDAVCTESGMP 240
DB 199 SKERKVCVEISCKSPDVINGSPIQKIYKENERPOYKCNMGYSEKSDAVCTESGMP 258
OY 241 LPSCGEKCDNPIYIPNGDYSPLRIRHGRGDELTTCRNGCFYPATGNTAKCTSTGWIPAP 300
|||||
```

```
DB 259 LPSCGEKCDNPIYIPNGDYSPLRIRHGRGDELTTCRNGCFYPATGNTAKCTSTGWIPAP 318
OY 301 RCTLKPDCDYPDIKKGHGLYHENMRPRFPV 329
|||||
DB 319 RCTLKPDCDYPDIKKGHGLYHENMRPRFPV 347

RESULT 3
BMSH
Complement factor H precursor - mouse
N:Alternate names: protein beta-1-H
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1987 #sequence, revision 30-Sep-1987 #text, change 22-Jun-1999
C:Accession: A26154; 149711; 149728
R:Kristensen, T.; Tack, B.F.
Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986
A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in lengt
A:Reference number: A26154; MVID:8623353
A:Accession: A26154
A:Molecule type: mRNA
A:Residues: 1-1234 <KRI>
A:Cross-references: GB:M12660; NID:q193724; PIDN:AAA37759.1; PID:q387181
R:Natsume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.
J. Immunol. 144, 358-362, 1990
A:Title: Demonstration of an unusual allelic variation of mouse factor H by the compl
A:Reference number: 149711; MVID:90111033
A:Accession: 149711
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-18 <RSS>
A:Cross-references: GB:M31979; NID:q193726; PIDN:AAA37762.1; PID:q193729
R:Funoz-Canoves, P.; Tack, B.F.; ViK, D.P.
Biochemistry 28, 9891-9897, 1989
A:Title: Analysis of complement factor H mRNA expression: Dexamethasone and IFN-gamma
A:Reference number: 149728; MVID:90148935
A:Accession: 149728
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-19 <R2>
A:Cross-references: GB:J02891; NID:q193805; PIDN:AAA37795.1; PID:q553926
C:Comment: Two codominant alleles of factor H are present in mice.
C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prot
(C5 convertase) in the alternative complement pathway.
C:Genetics:
A:Map position: 1
C:Superfamily: complement factor H; complement factor H repeat homology
C:Keywords: complement alternate pathway; duplication; glycoprotein; plasma
F:1-18/Domain: signal sequence #status predicted <StG>
F:19-1234/Product: complement factor H #status predicted <MPF>
F:21-80/Domain: complement factor H repeat homology <FH01>
F:85-141/Domain: complement factor H repeat homology <FH02>
F:146-205/Domain: complement factor H repeat homology <FH03>
F:210-262/Domain: complement factor H repeat homology <FH04>
F:246-248/Region: cell attachment (R-G-D) motif
F:267-320/Domain: complement factor H repeat homology <FH05>
F:325-385/Domain: complement factor H repeat homology <FH06>
F:389-442/Domain: complement factor H repeat homology <FH07>
F:448-505/Domain: complement factor H repeat homology <FH08>
F:509-564/Domain: complement factor H repeat homology <FH09>
F:569-622/Domain: complement factor H repeat homology <FH10>
F:629-683/Domain: complement factor H repeat homology <FH11>
F:690-743/Domain: complement factor H repeat homology <FH12>
F:752-802/Domain: complement factor H repeat homology <FH13>
F:808-861/Domain: complement factor H repeat homology <FH14>
F:867-931/Domain: complement factor H repeat homology <FH15>
F:936-989/Domain: complement factor H repeat homology <FH16>
F:994-1048/Domain: complement factor H repeat homology <FH17>
F:1053-1107/Domain: complement factor H repeat homology <FH18>
F:1114-1168/Domain: complement factor H repeat homology <FH19>
F:1172-1233/Domain: complement factor H repeat homology <FH20>
F:21-66:52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,
0-802,808-850,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,10
F:676,721,773,801,1030,1061,1225/Binding site: carbohydrate (Asn) (covalent) #status
```





C:Comment: The molecule has a central body supporting seven tentacles (alpha chains), c

C:Genetics:

A:Gene: GDB:C4BPA

A:Cross-references: GDB:120568; OMIM:120830

A:Map position: 1q32-1q32

A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3

C:Complex: octamer of seven alpha chains and one beta chain

C:Suprafamily: C4b-binding protein alpha chain; complement factor H repeat homology

C:Keywords: acute phase; chylomicron; complement pathway; duplication; glycoprotein; pla

F:1-48/Domain: signal sequence #status predicted <SIS>

F:49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>

F:50-108/Domain: complement factor H repeat homology <FH1>

F:113-170/Domain: complement factor H repeat homology <FH2>

F:175-234/Domain: complement factor H repeat homology <FH3>

F:239-294/Domain: complement factor H repeat homology <FH4>

F:299-360/Domain: complement factor H repeat homology <FH5>

F:364-422/Domain: complement factor H repeat homology <FH6>

F:381-404/Region: complement C4b binding #status predicted

F:426-480/Domain: complement factor H repeat homology <FH7>

F:484-538/Domain: complement factor H repeat homology <FH8>

F:521-506/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 18.8%; Score 353; DB 1; Length 597;

Best Local Similarity 29.3%; Pred. No. 4.8e-18;

Matches 100; Conservative 50; Mismatches 141; Indels 50; Gaps 21;

20 SDQTYEGTQAIYKCRPGY-RSLGNVIMVCR-GEWVALNPLKCRKRCRPGDPRPGT 77

119 --LRNQVEIKITDLSFGSQIEFSGSEGFLLIGSTSR-CEVQDRGVGWSHPLPQCEIVKC 175

129 LPTAENCGKIVSSAMPEDEYHFGQAVRVCNSGYKIEDEDEHHCDD---GFWSKEK 184

176 KPRPDRNRHSGE---ENFVAVGFSVTVSCDRPRLSLGHAISCEVENETIGVWRRSP 231

185 PKCEVSKSPDIVNGSPISQ-KIYKENERFOYKCMNGYSEKDAV--C-TESGWRP 240

232 PTECEIKTRKPRDYGSHGEMVSGFPIYNYKDTIYKCKQKQFVL--RGSSVTHCADSKWNP 289

241 L-PSCEKSCDN-PYIPNGDYS--PLRIK---HRTGDEITYOCNRNGYPAT-RGNTAKC- 291

290 SPSPACEPNSCTINLPDIPHASWETYPPTKEDYVGVTLNRCRPGYKPTTDEPTVICO 349

292 TSTGWIPAPRCTLKPCDYPDIKGGLYHENMRRP----YF 327

350 KNLRMWTPYOGCEALCCEPEPLNNGEITQHRKSRANHCYF 390

RESULT 8

complement C3b receptor type 2 long form precursor - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 08-Oct-1993

C:Accession: A45900; 148306

R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Weis, J.H.

J. Immunol. 144, 3581-3591, 1990

A:Title: The murine complement receptor gene family. IV. Alternative splicing of Cr2 ge

A:Reference number: A45900; M0ID:90229754

A:Accession: A45900

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-676 <KUR>

A:Cross-references: GB:M36470

A:Experimental sources: clone 31-1

R:Kim, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers, V.

J. Exp. Med. 181, 151-159, 1995

A:Title: Mouse complement regulatory protein Cr1/p65 uses the specific mechanisms of bc

A:Reference number: 148306; M0ID:95105691

A:Accession: 148306

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 21-367 <RES>

A:Cross-references: EMBL:U17128; NID:g9595980; PIDN:AAA78271.1; PID:g9595982

C:Genetics:

A:Gene: Cr2

A:Introns: 80/1; 113/2; 142/1; 275/1; 333/1

C:Suprafamily: complement factor H repeat homology

F:22-78/Domain: complement factor H repeat homology <FH01>

F:83-140/Domain: complement factor H repeat homology <FH02>

F:145-211/Domain: complement factor H repeat homology <FH03>

F:217-272/Domain: complement factor H repeat homology <FH04>

F:276-331/Domain: complement factor H repeat homology <FH05>

F:336-394/Domain: complement factor H repeat homology <FH06>

F:399-458/Domain: complement factor H repeat homology <FH07>

F:467-523/Domain: complement factor H repeat homology <FH08>

F:531-587/Domain: complement factor H repeat homology <FH09>

F:592-648/Domain: complement factor H repeat homology <FH10>

Query Match 18.8%; Score 353; DB 2; Length 676;

Best Local Similarity 28.9%; Pred. No. 5.5e-18;

Matches 93; Conservative 61; Mismatches 136; Indels 32; Gaps 17;

20 SDQTYEGTQAIYKCRPGY-RSLGNVIMVCR-GEWVALNPLKCRKRCRPGDPRPGT 78

37 SDKSEFALIGTWEYKCRPGYRFRKSFITTCLETSKW--SDAQCRKRCRPMNPQPLHGSV 94

79 TLNGVNEYGVAVVYTCNGCYOLLGEINREC---DTDGNNDIPICEVVKCLPYAP 134

95 HINTG--IEFGSTITVSCNGGYRLIGD-SSATGIVSNTVMMDMDPLCESIPCESPAI 151

135 ENKIVSSAMPEDEYHFGQAVRVCNSG-----YIIEDEDEHHCDD---GFWSKEK 164

152 SDCDFVSSRD---SFEYGVAVVYTCNGTGNREKLPDLVEKSTYCKSNQYGINMSP 208

185 PKCV-EISCKSPDIVNGSPISQ-KIYKENERFOYKCMNGYSEKDAV-C-TESGWRP 240

209 PCTIPRKYCKMPRELNLVSGFKHSFLLDYVIFKSKSFTMKGSIAMCOPRNSKSP 268

241 LPSCEKSCDNPYIPNGDYSPLRIKHTGDEITYOCNRNGYPATRGNTAKCTSTG-WI-P 298

269 LPFCFMCLPPONLHGDYNNKDEFFSVGQKVSYSVCNPQ-YTLIGTLVLCSTSLGTSMT 327

299 APRCTLKPCD-YPD-IKHGGLY 318

328 VPTCEVKSQDAIPNHLHGRVF 349

RESULT 9

complement C3b/C4b receptor, membrane-bound form precursor - human

N:Alternate names: complement C3b/C4b receptor; complement receptor type 1 (CR1); sur

N:Contains: complement C3b/C4b receptor, secreted form

C:Species: Homo sapiens (man)

C:Date: 02-Jul-1996 #sequence\_revision 24-Nov-1999 #text\_change 31-Jan-2000

C:Accession: I73012; 156203; A47602; S03291; S03843; A28507; A24748; C24748

R:Yik, D.P.; Wong, W.W.

J. Immunol. 151, 6214-6224, 1993

A:Title: Structure of the gene for the F allele of complement receptor type 1 and seq

A:Reference number: I56203; M0ID:94065175

A:Accession: I73012

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-683, 'X', 685-1133, 'X', 1135-1471, 'X', 1473-2489 <VIK1>

A:Cross-references: GB:U17418; NID:g306678; PID:g451303

A:Accession: I56203

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-683, 'X', 685-894, 'A', 896-1000, 1451-1471, 'X', 1473-2489 <VIK2>

A:Cross-references: GB:U17418; NID:g306678; PID:g306680

R:Wong, W.W.; Cahill, J.M.; Rosen, M.D.; Kennedy, C.A.; Bonaccio, E.T.; Morris, M.J.

J. Exp. Med. 169, 847-863, 1989



A:Title: Structure of the human CRI gene. Molecular basis of the structural and quantitative  
 A:Reference number: A47602; MUID:89176869  
 A:Accession: A47602  
 A:Molecule type: DNA  
 A:Residues: 1-41 <MON>  
 A:Cross-references: GB:X14893  
 R:Hourcade, D.; Miesner, D.R.; Atkinson, J.P.; Holers, V.M.  
 J. Exp. Med. 168, 1255-1270, 1988  
 A:Title: Identification of an alternative polyadenylation site in the human C3b/C4b receptor 1.  
 A:Reference number: S03291; MUID:89010527  
 A:Accession: S03291  
 A:Molecule type: mRNA  
 A:Residues: 26-584 <HOU>  
 A:Cross-references: EMBL:X14362; NID:930197; PID:9736240  
 A:Experimental source: Clone CRI 4  
 R:Klickstein, L.B.; Bartow, T.J.; Miletic, V.; Rabson, L.D.; Smith, J.A.; Fearon, D.T.  
 J. Exp. Med. 168, 1699-1717, 1988  
 A:Title: Identification of distinct C3b and C4b recognition sites in the human C3b/C4b receptor.  
 A:Reference number: S03843; MUID:89035992  
 A:Accession: S03843  
 A>Status: translation not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-894, 'A', 896-1000, 1451-2064, 'T', 2066-2276, 'P', 2278-2299, 'H', 2301-2325, 'T', 2327-2332, 'A', 2333-2334, 'G', 2335-2336, 'C', 2337-2338, 'G', 2339-2340, 'A', 2341-2342, 'G', 2343-2344, 'A', 2345-2346, 'G', 2347-2348, 'A', 2349-2350, 'G', 2351-2352, 'A', 2353-2354, 'G', 2355-2356, 'A', 2357-2358, 'G', 2359-2360, 'A', 2361-2362, 'G', 2363-2364, 'A', 2365-2366, 'G', 2367-2368, 'A', 2369-2370, 'G', 2371-2372, 'A', 2373-2374, 'G', 2375-2376, 'A', 2377-2378, 'G', 2379-2380, 'A', 2381-2382, 'G', 2383-2384, 'A', 2385-2386, 'G', 2387-2388, 'A', 2389-2390, 'G', 2391-2392, 'A', 2393-2394, 'G', 2395-2396, 'A', 2397-2398, 'G', 2399-2400, 'A', 2401-2402, 'G', 2403-2404, 'A', 2405-2406, 'G', 2407-2408, 'A', 2409-2410, 'G', 2411-2412, 'A', 2413-2414, 'G', 2415-2416, 'A', 2417-2418, 'G', 2419-2420, 'A', 2421-2422, 'G', 2423-2424, 'A', 2425-2426, 'G', 2427-2428, 'A', 2429-2430, 'G', 2431-2432, 'A', 2433-2434, 'G', 2435-2436, 'A', 2437-2438, 'G', 2439-2440, 'A', 2441-2442, 'G', 2443-2444, 'A', 2445-2446, 'G', 2447-2448, 'A', 2449-2450, 'G', 2451-2452, 'A', 2453-2454, 'G', 2455-2456, 'A', 2457-2458, 'G', 2459-2460, 'A', 2461-2462, 'G', 2463-2464, 'A', 2465-2466, 'G', 2467-2468, 'A', 2469-2470, 'G', 2471-2472, 'A', 2473-2474, 'G', 2475-2476, 'A', 2477-2478, 'G', 2479-2480, 'A', 2481-2482, 'G', 2483-2484, 'A', 2485-2486, 'G', 2487-2488, 'A', 2489-2490, 'G', 2491-2492, 'A', 2493-2494, 'G', 2495-2496, 'A', 2497-2498, 'G', 2499-2500, 'A', 2501-2502, 'G', 2503-2504, 'A', 2505-2506, 'G', 2507-2508, 'A', 2509-2510, 'G', 2511-2512, 'A', 2513-2514, 'G', 2515-2516, 'A', 2517-2518, 'G', 2519-2520, 'A', 2521-2522, 'G', 2523-2524, 'A', 2525-2526, 'G', 2527-2528, 'A', 2529-2530, 'G', 2531-2532, 'A', 2533-2534, 'G', 2535-2536, 'A', 2537-2538, 'G', 2539-2540, 'A', 2541-2542, 'G', 2543-2544, 'A', 2545-2546, 'G', 2547-2548, 'A', 2549-2550, 'G', 2551-2552, 'A', 2553-2554, 'G', 2555-2556, 'A', 2557-2558, 'G', 2559-2560, 'A', 2561-2562, 'G', 2563-2564, 'A', 2565-2566, 'G', 2567-2568, 'A', 2569-2570, 'G', 2571-2572, 'A', 2573-2574, 'G', 2575-2576, 'A', 2577-2578, 'G', 2579-2580, 'A', 2581-2582, 'G', 2583-2584, 'A', 2585-2586, 'G', 2587-2588, 'A', 2589-2590, 'G', 2591-2592, 'A', 2593-2594, 'G', 2595-2596, 'A', 2597-2598, 'G', 2599-2600, 'A', 2601-2602, 'G', 2603-2604, 'A', 2605-2606, 'G', 2607-2608, 'A', 2609-2610, 'G', 2611-2612, 'A', 2613-2614, 'G', 2615-2616, 'A', 2617-2618, 'G', 2619-2620, 'A', 2621-2622, 'G', 2623-2624, 'A', 2625-2626, 'G', 2627-2628, 'A', 2629-2630, 'G', 2631-2632, 'A', 2633-2634, 'G', 2635-2636, 'A', 2637-2638, 'G', 2639-2640, 'A', 2641-2642, 'G', 2643-2644, 'A', 2645-2646, 'G', 2647-2648, 'A', 2649-2650, 'G', 2651-2652, 'A', 2653-2654, 'G', 2655-2656, 'A', 2657-2658, 'G', 2659-2660, 'A', 2661-2662, 'G', 2663-2664, 'A', 2665-2666, 'G', 2667-2668, 'A', 2669-2670, 'G', 2671-2672, 'A', 2673-2674, 'G', 2675-2676, 'A', 2677-2678, 'G', 2679-2680, 'A', 2681-2682, 'G', 2683-2684, 'A', 2685-2686, 'G', 2687-2688, 'A', 2689-2690, 'G', 2691-2692, 'A', 2693-2694, 'G', 2695-2696, 'A', 2697-2698, 'G', 2699-2700, 'A', 2701-2702, 'G', 2703-2704, 'A', 2705-2706, 'G', 2707-2708, 'A', 2709-2710, 'G', 2711-2712, 'A', 2713-2714, 'G', 2715-2716, 'A', 2717-2718, 'G', 2719-2720, 'A', 2721-2722, 'G', 2723-2724, 'A', 2725-2726, 'G', 2727-2728, 'A', 2729-2730, 'G', 2731-2732, 'A', 2733-2734, 'G', 2735-2736, 'A', 2737-2738, 'G', 2739-2740, 'A', 2741-2742, 'G', 2743-2744, 'A', 2745-2746, 'G', 2747-2748, 'A', 2749-2750, 'G', 2751-2752, 'A', 2753-2754, 'G', 2755-2756, 'A', 2757-2758, 'G', 2759-2760, 'A', 2761-2762, 'G', 2763-2764, 'A', 2765-2766, 'G', 2767-2768, 'A', 2769-2770, 'G', 2771-2772, 'A', 2773-2774, 'G', 2775-2776, 'A', 2777-2778, 'G', 2779-2780, 'A', 2781-2782, 'G', 2783-2784, 'A', 2785-2786, 'G', 2787-2788, 'A', 2789-2790, 'G', 2791-2792, 'A', 2793-2794, 'G', 2795-2796, 'A', 2797-2798, 'G', 2799-2800, 'A', 2801-2802, 'G', 2803-2804, 'A', 2805-2806, 'G', 2807-2808, 'A', 2809-2810, 'G', 2811-2812, 'A', 2813-2814, 'G', 2815-2816, 'A', 2817-2818, 'G', 2819-2820, 'A', 2821-2822, 'G', 2823-2824, 'A', 2825-2826, 'G', 2827-2828, 'A', 2829-2830, 'G', 2831-2832, 'A', 2833-2834, 'G', 2835-2836, 'A', 2837-2838, 'G', 2839-2840, 'A', 2841-2842, 'G', 2843-2844, 'A', 2845-2846, 'G', 2847-2848, 'A', 2849-2850, 'G', 2851-2852, 'A', 2853-2854, 'G', 2855-2856, 'A', 2857-2858, 'G', 2859-2860, 'A', 2861-2862, 'G', 2863-2864, 'A', 2865-2866, 'G', 2867-2868, 'A', 2869-2870, 'G', 2871-2872, 'A', 2873-2874, 'G', 2875-2876, 'A', 2877-2878, 'G', 2879-2880, 'A', 2881-2882, 'G', 2883-2884, 'A', 2885-2886, 'G', 2887-2888, 'A', 2889-2890, 'G', 2891-2892, 'A', 2893-2894, 'G', 2895-2896, 'A', 2897-2898, 'G', 2899-2900, 'A', 2901-2902, 'G', 2903-2904, 'A', 2905-2906, 'G', 2907-2908, 'A', 2909-2910, 'G', 2911-2912, 'A', 2913-2914, 'G', 2915-2916, 'A', 2917-2918, 'G', 2919-2920, 'A', 2921-2922, 'G', 2923-2924, 'A', 2925-2926, 'G', 2927-2928, 'A', 2929-2930, 'G', 2931-2932, 'A', 2933-2934, 'G', 2935-2936, 'A', 2937-2938, 'G', 2939-2940, 'A', 2941-2942, 'G', 2943-2944, 'A', 2945-2946, 'G', 2947-2948, 'A', 2949-2950, 'G', 2951-2952, 'A', 2953-2954, 'G', 2955-2956, 'A', 2957-2958, 'G', 2959-2960, 'A', 2961-2962, 'G', 2963-2964, 'A', 2965-2966, 'G', 2967-2968, 'A', 2969-2970, 'G', 2971-2972, 'A', 2973-2974, 'G', 2975-2976, 'A', 2977-2978, 'G', 2979-2980, 'A', 2981-2982, 'G', 2983-2984, 'A', 2985-2986, 'G', 2987-2988, 'A', 2989-2990, 'G', 2991-2992, 'A', 2993-2994, 'G', 2995-2996, 'A', 2997-2998, 'G', 2999-3000, 'A', 3001-3002, 'G', 3003-3004, 'A', 3005-3006, 'G', 3007-3008, 'A', 3009-3010, 'G', 3011-3012, 'A', 3013-3014, 'G', 3015-3016, 'A', 3017-3018, 'G', 3019-3020, 'A', 3021-3022, 'G', 3023-3024, 'A', 3025-3026, 'G', 3027-3028, 'A', 3029-3030, 'G', 3031-3032, 'A', 3033-3034, 'G', 3035-3036, 'A', 3037-3038, 'G', 3039-3040, 'A', 3041-3042, 'G', 3043-3044, 'A', 3045-3046, 'G', 3047-3048, 'A', 3049-3050, 'G', 3051-3052, 'A', 3053-3054, 'G', 3055-3056, 'A', 3057-3058, 'G', 3059-3060, 'A', 3061-3062, 'G', 3063-3064, 'A', 3065-3066, 'G', 3067-3068, 'A', 3069-3070, 'G', 3071-3072, 'A', 3073-3074, 'G', 3075-3076, 'A', 3077-3078, 'G', 3079-3080, 'A', 3081-3082, 'G', 3083-3084, 'A', 3085-3086, 'G', 3087-3088, 'A', 3089-3090, 'G', 3091-3092, 'A', 3093-3094, 'G', 3095-3096, 'A', 3097-3098, 'G', 3099-3100, 'A', 3101-3102, 'G', 3103-3104, 'A', 3105-3106, 'G', 3107-3108, 'A', 3109-3110, 'G', 3111-3112, 'A', 3113-3114, 'G', 3115-3116, 'A', 3117-3118, 'G', 3119-3120, 'A', 3121-3122, 'G', 3123-3124, 'A', 3125-3126, 'G', 3127-3128, 'A', 3129-3130, 'G', 3131-3132, 'A', 3133-3134, 'G', 3135-3136, 'A', 3137-3138, 'G', 3139-3140, 'A', 3141-3142, 'G', 3143-3144, 'A', 3145-3146, 'G', 3147-3148, 'A', 3149-3150, 'G', 3151-3152, 'A', 3153-3154, 'G', 3155-3156, 'A', 3157-3158, 'G', 3159-3160, 'A', 3161-3162, 'G', 3163-3164, 'A', 3165-3166, 'G', 3167-3168, 'A', 3169-3170, 'G', 3171-3172, 'A', 3173-3174, 'G', 3175-3176, 'A', 3177-3178, 'G', 3179-3180, 'A', 3181-3182, 'G', 3183-3184, 'A', 3185-3186, 'G', 3187-3188, 'A', 3189-3190, 'G', 3191-3192, 'A', 3193-3194, 'G', 3195-3196, 'A', 3197-3198, 'G', 3199-3200, 'A', 3201-3202, 'G', 3203-3204, 'A', 3205-3206, 'G', 3207-3208, 'A', 3209-3210, 'G', 3211-3212, 'A', 3213-3214, 'G', 3215-3216, 'A', 3217-3218, 'G', 3219-3220, 'A', 3221-3222, 'G', 3223-3224, 'A', 3225-3226, 'G', 3227-3228, 'A', 3229-3230, 'G', 3231-3232, 'A', 3233-3234, 'G', 3235-3236, 'A', 3237-3238, 'G', 3239-3240, 'A', 3241-3242, 'G', 3243-3244, 'A', 3245-3246, 'G', 3247-3248, 'A', 3249-3250, 'G', 3251-3252, 'A', 3253-3254, 'G', 3255-3256, 'A', 3257-3258, 'G', 3259-3260, 'A', 3261-3262, 'G', 3263-3264, 'A', 3265-3266, 'G', 3267-3268, 'A', 3269-3270, 'G', 3271-3272, 'A', 3273-3274, 'G', 3275-3276, 'A', 3277-3278, 'G', 3279-3280, 'A', 3281-3282, 'G', 3283-3284, 'A', 3285-3286, 'G', 3287-3288, 'A', 3289-3290, 'G', 3291-3292, 'A', 3293-3294, 'G', 3295-3296, 'A', 3297-3298, 'G', 3299-3300, 'A', 3301-3302, 'G', 3303-3304, 'A', 3305-3306, 'G', 3307-3308, 'A', 3309-3310, 'G', 3311-3312, 'A', 3313-3314, 'G', 3315-3316, 'A', 3317-3318, 'G', 3319-3320, 'A', 3321-3322, 'G', 3323-3324, 'A', 3325-3326, 'G', 3327-3328, 'A', 3329-3330, 'G', 3331-3332, 'A', 3333-3334, 'G', 3335-3336, 'A', 3337-3338, 'G', 3339-3340, 'A', 3341-3342, 'G', 3343-3344, 'A', 3345-3346, 'G', 3347-3348, 'A', 3349-3350, 'G', 3351-3352, 'A', 3353-3354, 'G', 3355-3356, 'A', 3357-3358, 'G', 3359-3360, 'A', 3361-3362, 'G', 3363-3364, 'A', 3365-3366, 'G', 3367-3368, 'A', 3369-3370, 'G', 3371-3372, 'A', 3373-3374, 'G', 3375-3376, 'A', 3377-3378, 'G', 3379-3380, 'A', 3381-3382, 'G', 3383-3384, 'A', 3385-3386, 'G', 3387-3388, 'A', 3389-3390, 'G', 3391-3392, 'A', 3393-3394, 'G', 3395-3396, 'A', 3397-3398, 'G', 3399-3400, 'A', 3401-3402, 'G', 3403-3404, 'A', 3405-3406, 'G', 3407-3408, 'A', 3409-3410, 'G', 3411-3412, 'A', 3413-3414, 'G', 3415-3416, 'A', 3417-3418, 'G', 3419-3420, 'A', 3421-3422, 'G', 3423-3424, 'A', 3425-3426, 'G', 3427-3428, 'A', 3429-3430, 'G', 3431-3432, 'A', 3433-3434, 'G', 3435-3436, 'A', 3437-3438, 'G', 3439-3440, 'A', 3441-3442, 'G', 3443-3444, 'A', 3445-3446, 'G', 3447-3448, 'A', 3449-3450, 'G', 3451-3452, 'A', 3453-3454, 'G', 3455-3456, 'A', 3457-3458, 'G', 3459-3460, 'A', 3461-3462, 'G', 3463-3464, 'A', 3465-3466, 'G', 3467-3468, 'A', 3469-3470, 'G', 3471-3472, 'A', 3473-3474, 'G', 3475-3476, 'A', 3477-3478, 'G', 3479-3480, 'A', 3481-3482, 'G', 3483-3484, 'A', 3485-3486, 'G', 3487-3488, 'A', 3489-3490, 'G', 3491-3492, 'A', 3493-3494, 'G', 3495-3496, 'A', 3497-3498, 'G', 3499-3500, 'A', 3501-3502, 'G', 3503-3504, 'A', 3505-3506, 'G', 3507-3508, 'A', 3509-3510, 'G', 3511-3512, 'A', 3513-3514, 'G', 3515-3516, 'A', 3517-3518, 'G', 3519-3520, 'A', 3521-3522, 'G', 3523-3524, 'A', 3525-3526, 'G', 3527-3528, 'A', 3529-3530, 'G', 3531-3532, 'A', 3533-3534, 'G', 3535-3536, 'A', 3537-3538, 'G', 3539-3540, 'A', 3541-3542, 'G', 3543-3544, 'A', 3545-3546, 'G', 3547-3548, 'A', 3549-3550, 'G', 3551-3552, 'A', 3553-3554, 'G', 3555-3556, 'A', 3557-3558, 'G', 3559-3560, 'A', 3561-3562, 'G', 3563-3564, 'A', 3565-3566, 'G', 3567-3568, 'A', 3569-3570, 'G', 3571-3572, 'A', 3573-3574, 'G', 3575-3576, 'A', 3577-3578, 'G', 3579-3580, 'A', 3581-3582, 'G', 3583-3584, 'A', 3585-3586, 'G', 3587-3588, 'A', 3589-3590, 'G', 3591-3592, 'A', 3593-3594, 'G', 3595-3596, 'A', 3597-3598, 'G', 3599-3600, 'A', 3601-3602, 'G', 3603-3604, 'A', 3605-3606, 'G', 3607-3608, 'A', 3609-3610, 'G', 3611-3612, 'A', 3613-3614, 'G', 3615-3616, 'A', 3617-3618, 'G', 3619-3620, 'A', 3621-3622, 'G', 3623-3624, 'A', 3625-3626, 'G', 3627-3628, 'A', 3629-3630, 'G', 3631-3632, 'A', 3633-3634, 'G', 3635-3636, 'A', 3637-3638, 'G', 3639-3640, 'A', 3641-3642, 'G', 3643-3644, 'A', 3645-3646, 'G', 3647-3648, 'A', 3649-3650, 'G', 3651-3652, 'A', 3653-3654, 'G', 3655-3656, 'A', 3657-3658, 'G', 3659-3660, 'A', 3661-3662, 'G', 3663-3664, 'A', 3665-3666, 'G', 3667-3668, 'A', 3669-3670, 'G', 3671-3672, 'A', 3673-3674, 'G', 3675-3676, 'A', 3677-3678, 'G', 3679-3680, 'A', 3681-3682, 'G', 3683-3684, 'A', 3685-3686, 'G', 3687-3688, 'A', 3689-3690, 'G', 3691-3692, 'A', 3693-3694, 'G', 3695-3696, 'A', 3697-3698, 'G', 3699-3700, 'A', 3701-3702, 'A', 3703-3704, 'G', 3705-3706, 'A', 3707-3708, 'G', 3709-3710, 'A', 3711-3712, 'G', 3713-3714, 'A', 3715-3716, 'G', 3717-3718, 'A', 3719-3720, 'G', 3721-3722, 'A', 3723-3724, 'G', 3725-3726, 'A', 3727-3728, 'G', 3729-3730, 'A', 3731-3732, 'G', 3733-3734, 'A', 3735-3736, 'G', 3737-3738, 'A', 3739-3740, 'G', 3741-3742, 'A', 3743-3744, 'G', 3745-3746, 'A', 3747-3748, 'G', 3749-3750, 'A', 3751-3752, 'G', 3753-3754, 'A', 3755-3756, 'G', 3757-3758, 'A', 3759-3760, 'G', 3761-3762, 'A', 3763-3764, 'G', 3765-3766, 'A', 3767-3768, 'A', 3769-3770, 'G', 3771-3772, 'A', 3773-3774, 'G', 3775-3776, 'A', 3777-3778, 'G', 3779-3780, 'A', 3781-3782, 'G', 3783-3784, 'A', 3785-3786, 'G', 3787-3788, 'A', 3789-3790, 'G', 3791-3792, 'A', 3793-3794, 'G', 3795-3796, 'A', 3797-3798, 'G', 3799-3800, 'A', 3801-3802, 'G', 3803-3804, 'A', 3805-3806, 'G', 3807-3808, 'A', 3809-3810, 'G', 3811-3812, 'A', 3813-3814, 'G', 3815-3816, 'A', 3817-3818, 'G', 3819-3820, 'A', 3821-3822, 'G', 3823-3824, 'A', 3825-3826, 'G', 3827-3828, 'A', 3829-3830, 'G', 3831-3832, 'A', 3833-3834, 'G', 3835-3836, 'A', 3837-3838, 'G', 3839-3840, 'A', 3841-3842, 'G', 3843-3844, 'A', 3845-3846, 'G', 3847-3848, 'A', 3849-3850, 'G', 3851-3852, 'A', 3853-3854, 'G', 3855-3856, 'A', 3857-3858, 'G', 3859-3860, 'A', 3861-3862, 'G', 3863-3864, 'A', 3865-3866, 'G', 3867-3868, 'A', 3869-3870, 'G', 3871-3872, 'A', 3873-3874, 'G', 3875-3876, 'A', 3877-3878, 'G', 3879-3880, 'A', 3881-3882, 'G', 3883-3884, 'A', 3885-3886, 'G', 3887-3888, 'A', 3889-3890, 'G', 3891-3892, 'A', 3893-3894, 'G', 3895-3896, 'A', 3897-3898, 'G', 3899-3900, 'A', 3901-3902, 'G', 3903-3904, 'A', 3905-3906, 'G', 3907-3908, 'A', 3909-3910, 'G', 3911-3912, 'A', 3913-3914, 'G', 3915-3916, 'A', 3917-3918, 'G', 3919-3920, 'A', 3921-3922, 'G', 3923-3924, 'A', 3925-3926, 'G', 3927-3928, 'A', 3929-3930, 'G', 3931-3932, 'A', 3933-3934, 'G', 3935-3936, 'A', 3937-3938, 'G', 3939-3940, 'A', 3941-3942, 'G', 3943-3944, 'A', 3945-3946, 'G', 3947-3948, 'A', 3949-3950, 'G', 3951-3952, 'A', 3953-3954, 'G', 3955-3956, 'A', 3957-3958, 'G', 3959-3960, 'A', 3961-3962, 'G', 3963-3964, 'A', 3965-3966, 'G', 3967-3968, 'A', 3969-3970, 'G', 3971-3972, 'A', 3973-3974, 'G', 3975-3976, 'A', 3977-3978, 'G', 3979-3980, 'A', 3981-3982, 'G', 3983-3984, 'A', 3985-3986, 'G', 3987-3988, 'A', 3989-3990, 'G', 3991-3992, 'A', 3993-3994,





QY 69 HPGDPTGFTLTGNGVEYKAVYTCNBSXOLLGELINTECDTGD-WTNDIPICEYVK 127  
 Db 133 DRGIENG---LRGDDFEIPIHHVYKSCNPFLLVSTG-RQCSNMGEMTEPAPACKATE 188  
 QY 128 CLPYTAPENGKIVSSAMEPDRREYHFGQAVRFVCSGYKIEGDEMHGCDDEGFWSEKRC 187  
 Db 189 CSRPSSPLHGKVVSSSLT-----YQSVVYVSCDHGRVLGQVORICLAEIKMGNEPRC 242  
 QY 188 VEISCK-SPIVING-----SPT----- 203  
 Db 243 EEIRCSVLPFLPNGLYIGSETSGAVAVFRCLTETHEGASKAKCMEDQWSPAPICLA 302  
 QY 204 SOKIYKENERFOYK-----CNMGXEYSEKADACTESGWRPLPSCERKS 248  
 Db 303 SCRVPHIQNKIRKSSGQLASSKATVECNKOEANIDERLISNSTWSHVPVCSPLS 362  
 QY 249 CDN--PYIPNGDYSPLRIKHTGDEITYQCRNGFPYATRGNTAKTSTGWT---PAPRCT 303  
 Db 363 CHWMPRPVPHARI--LFSKSSHGSIAKYECNNGYHPRNNOIILCLGEMTKDGPMPKCL 420  
 QY 304 LKPCDYPDIDHGGI 317  
 Db 421 PSCWDEHPSKTYGTL 434

# RESULT 13

T42921

complement control protein homolog ccph - ateline herpesvirus 3 (strain 73)  
 C:Species: ateline herpesvirus 3  
 A:Variety: strain 73  
 C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 04-Mar-2000  
 A:Accession: T42921  
 R:Albrecht, J.C.: Fleckenstein, B.  
 A:Description: Primary structure of the herpesvirus ateles genome.  
 A:Reference number: Z22274  
 A:Accession: T42921  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-360 <ALB>  
 A:Cross-references: EMBL:AF083424; PIDN:MAC95530.1  
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog

Query Match 17.9%; Score 335; DB 2; Length 360;  
 Best Local Similarity 33.4%; Pred. No. 5.5e-17;  
 Matches 103; Conservative 32; Mismatches 119; Indels 54; Gaps 20;

QY 8 PRNRTEI-----LNGSMQDTYPEGTQAIYKCRPGYRSLGNVT--MVCRRKGWVALNPL 59  
 Db 24 PRNRVYSLRYVNITN--SSGSYPNCTITQVYCRKCY--IGKIQIVTYCVNGWVTPN-- 77  
 QY 60 RRCQKRPCHPGDTPPGTTLTGAVNEFYKAVYTCNBSXOLLGELINTECDTGD--DTG-- 115  
 Db 78 -BCORRRCSTPADLNGWYTVT-GNLY-YGSVITYTCNGLYQLGSP--SSCLGPDGRV 133  
 QY 116 -WTNIDIPICEYVKCLPYTAPENGKIVSSAMEPDRREYH-FGQAVRFVCSNGYK--IEGDEE 171  
 Db 134 NMTTPRPICEITKCKRPPPIANGTHNT-----KEYITVLAVYVASCNDETKTLTLGPS 188  
 QY 172 MHCSDDGFW-SREKPRCWEISCKSPDVINGSPISOKIYKENERFOY---KCNMGXEYSE 227  
 Db 189 KOCSETGRVNPDEETKCEKFKVCKIPOVANGHEVKT--SNVQYQYVAKIKCDKGRLOG 246  
 QY 228 RCDVACTESGWRP-LPSCERKSCNDPYIPNGDYSPLRIKHTGDEITYQCRNGFPYATRG 286  
 Db 247 ETPNCKNGWMPALPTCK-----PAPRGD-----MPHIDSGEDTST-----PSGRN 290  
 QY 287 NTAKTST 294  
 Db 291 CNQCTTS 298

# RESULT 14

JC2054

complement regulatory protein, 512 antigen precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 21-Jan-2000  
 A:Accession: JC2054; PC2027  
 R:Sakurada, C.; Seno, H.; Dohl, N.; Takizawa, H.; Nonaka, M.; Okada, N.; Okada, H.  
 A:Title: Molecular cloning of the rat complement regulatory protein, 512 antigen.  
 A:Reference number: JC2054; MUID:94161746  
 A:Accession: JC2054  
 A:Molecule type: mRNA  
 A:Residues: 1-497 <SAK>  
 A:Accession: PC2027  
 A:Molecule type: protein  
 A:Residues: 39-51 <SN2>

C:Comment: This protein plays a critical role in protection against complement mediated  
 C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 C:Keywords: glycoprotein; transmembrane protein  
 F:1-36/Domain: signal sequence #status predicted <SIG>  
 F:37-488/Product: complement regulatory protein, 512 antigen #status predicted <MAT>  
 F:38-94/Domain: complement factor H repeat homology <FH1>  
 F:98-156/Domain: complement factor H repeat homology <FH2>  
 F:161-227/Domain: complement factor H repeat homology <FH3>  
 F:233-288/Domain: complement factor H repeat homology <FH4>  
 F:294-352/Domain: complement factor H repeat homology <FH5>  
 F:357-413/Domain: complement factor H repeat homology <FH6>  
 F:421-444/Domain: transmembrane #status predicted <TM>  
 F:247,331,346,450,482,483/Binding site: carboxylate (Asn) (covalent) #status predict

Query Match 17.2%; Score 323; DB 2; Length 497;  
 Best Local Similarity 28.3%; Pred. No. 5.6e-16;  
 Matches 95; Conservative 55; Mismatches 144; Indels 42; Gaps 21;

QY 21 DQYTPETQAIYKCRPGYRSLGNVIMCRKGEWVALNPLRCKORPCGHGDPPTGFTL 80  
 Db 55 ESTPVPSTLSKYECPYIKRPSITCEVNSVMT--SPQDVCIKQCEPLDPQNGIVH 112  
 QY 81 TGCNVEFYKAVYTCNBSXOLLGELINTECDTGD--GWNIDIPICEYVK-LPYTPEN 136  
 Db 113 --NTDIRFGSSITTYTCNEGTRLIGSSAMCTISDQSYAWAEAPICISICEIPSPIS 169  
 QY 137 GKIYSSAMEPDR-YHFGQAVRFVCSNG-----YKIEGDEMHGCS---DDGFWSEKXP 185  
 Db 170 GDFFS-----PNRDEPHYGMVVTYQCNLDARGKFLNLYGEPHSIHCTSIDQGVWMSGPPP 225  
 QY 186 KCEYIS-CKSPDVINGSPISQ-KIYKENERFOYKCNMGYEYSEKGDVCTE-SGWRP-L 241  
 Db 226 QCIEELMKCTPPHENVAVTYSKNSKLSFLRDMVEFRQDGMKGDSSVYCRSLNRMPEOL 285  
 QY 242 PSC-EKESCDN--PYIPNGD-YSPRLKHTGDEITYQCRNGFPYATRGNTAK-CTSTG- 295  
 Db 286 PSCFKYKSCGAFGLPENGHVFVPOU--OLGAKVTFVCTGY--OLKGNSSSHCVLDGV 341  
 QY 296 --W-IPAPRCTLKPDYDPIDIKHGLYHNMRRPYE 327  
 Db 342 ESTMNSVPCVEQVICKLPQDMSGFQGLQMKKDY 377

# RESULT 15

I46001

C4b-binding protein alpha chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 A:Accession: I46001; S43190  
 R:Hillarp, A.; Thern, A.; Dahlback, B.  
 J:Immunol. 153, 4190-4199, 1994  
 A:Title: Bovine C4b binding protein. Molecular cloning of the alpha- and beta-chains  
 A:Reference number: I46001; MUID:95015909  
 A:Accession: I46001



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2000, 16:47:33 ; Search time 47.7 seconds  
(without alignments)  
352.557 Million cell updates/sec

Title: US-09-316-163-10

Perfect score: 1497

Sequence: 1 EDCNLEPPRRNTEILTGSWS.....EKSCDPYIPNGDYSLRLIK 265

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

1: PIR.65:\*  
2: pirl:\*  
3: pirl:\*  
4: pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1497	100.0	449	1 NBHHS	complement factor
2	1497	100.0	1231	1 NBHHS	complement factor
3	1024	68.4	1234	1 NBHHS	complement factor
4	699	46.7	669	2 S6551	factor H - bovine
5	440	29.4	1053	2 S46199	probable complemen
6	332.5	22.2	360	2 T42921	complement control
7	303	20.2	597	1 NBHUC4	C4b-binding protel
8	302	20.2	597	1 S53711	C4BP alpha chain p
9	301.5	20.1	263	1 WMYZSP	apolipoprotein H h
10	301	20.1	676	2 A45900	complement C3b rec
11	298.5	19.9	2489	2 I73012	complement C3b/C4b
12	290.5	19.4	2014	2 I36936	complement recepto
13	289.5	19.3	497	2 JC2054	complement regulat
14	284.5	19.0	263	1 C36838	hypothetical prote
15	284.5	19.0	263	2 T28450	hypothetical prote
16	284.5	19.0	482	2 A34924	complement C3b/C4b
17	283.5	18.9	263	2 B72152	B18L protein - var
18	277.5	18.5	560	2 T16833	hypothetical prote
19	276.5	18.5	558	2 S57953	C4BP protein alpha
20	275	18.4	469	1 NBHSC4	C4b-binding protel
21	274.5	18.3	349	2 G02813	sperm CD46 - human
22	274.5	18.3	369	2 I57998	membrane cofactor
23	274	18.3	377	2 I54479	membrane cofactor
24	274	18.3	384	2 S01896	membrane cofactor
25	273.5	18.3	433	2 A30550	complement C3b/C4b
26	272	18.2	610	2 A35046	E-selectin precurs
27	270.5	18.1	362	2 JC5194	membrane cofactor
28	270.5	18.1	369	2 JC5138	membrane cofactor
29	270.5	18.1	440	2 A43519	complement recepto

30	266	17.8	482	2 JC5092	E-selectin - pig
31	265.5	17.7	610	1 T46001	C4b-binding protel
32	264.5	17.7	302	1 WMBE1E	secretory compleme
33	264.5	17.7	360	1 WMBE2E	membrane-bound com
34	262	17.5	381	1 B26359	decay-accelerating
35	262	17.5	440	2 A26359	decay-accelerating
36	260.5	17.4	768	2 A42755	P-selectin precurs
37	260	17.4	612	2 S23174	endothelial leukoc
38	257.5	17.2	1025	1 A43526	complement C3d/Eps
39	257	17.2	579	2 A56740	sperm-egg recognit
40	256.5	17.1	618	2 B42755	E-selectin precurs
41	255	17.0	452	2 A35068	complement factor
42	254	17.0	768	2 I53821	P-selectin - rat
43	253	16.9	345	1 NBMS	apolipoprotein H p
44	252.5	16.9	1091	1 PL0009	complement C3d/Eps
45	252	16.8	340	2 I56234	decay-accelerating

#### ALIGNMENTS

RESULT 1  
NBHHS  
complement factor H precursor, short splice form - human  
N:Alternate names: complement factor H-related protein; complement protein H  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence-revision 23-Feb-1996 #text-change 22-Jun-1999  
C:Accession: S03013; B60238; A27877; A61103; A26505; S10479  
R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
Biochem. J. 249, 593-602, 1988  
A:Title: The complete amino acid sequence of human complement factor H.  
A:Reference number: S00254; MUID:88134059  
A:Accession: S03013  
A:Molecule type: mRNA  
A:Residues: 1-449 <RIP>  
A:Cross-references: EMBL:X07523; EMBL:Y00716; NID:932492; PIDN:CAA30403.1; PID:975807  
A:Note: part of this sequence, including the amino end of the mature protein was conf  
A:Note: 402-Tyr was also found  
R:Estalier, C.; Schwaible, W.; Dierich, M.; Weiss, E.H.  
Eur. J. Immunol. 21, 799-802, 1991  
A:Title: Human complement factor H: two factor H proteins are derived from alternativ  
A:Reference number: A60238; MUID:91184292  
A:Accession: B60238  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-33;434-449 <EST>  
A:Note: only portions of this 1.8 kilobase mRNA were sequenced  
R:Schulz, T.F.; Schwaible, W.; Stanley, K.K.; Weiss, E.; Dierich, M.P.  
Eur. J. Immunol. 16, 1351-1355, 1986  
A:Title: Human complement factor H: isolation of cDNA clones and partial cDNA sequence  
A:Reference number: A27877; MUID:87054207  
A:Accession: A27877  
A:Molecule type: mRNA  
A:Residues: 1-55;401,'Y',403-449 <SCH>  
A:Note: an additional nucleotide present within the codon for Glu-310 was thought to  
R:Schwaible, W.; Zwirner, J.; Schulz, T.F.; Linke, R.P.; Dierich, M.P.; Weiss, E.H.  
Eur. J. Immunol. 17, 1485-1489, 1987  
A:Title: Human complement factor H: expression of an additional truncated gene produc  
A:Reference number: A61103; MUID:88055295  
A:Accession: A61103  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 27-76 <SC2>  
A:Note: this is a partial sequence of an alternatively spliced 1.8 kilobase mRNA that  
R:Sim, R.B.; Discipio, R.G.  
Biochem. J. 205, 285-293, 1982  
A:Title: Purification and structural studies on the complement-system control protein  
A:Reference number: A26505; MUID:83048213  
A:Accession: A26505  
A:Molecule type: protein  
A:Residues: 19-20,'Q',22-29,'V',31-33,'Q',35 <SIM>  
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Driscoll, P  
Biochemistry 31, 3626-3634, 1992

A:Title: Solution structure of the fifth repeat of factor H: A second example of the com  
A:Reference number: A44551; MUID:92232649  
A:Contents: annotation: NMR structure determination, residues 264-292  
R:Kristensen, T.; Metseel, R.A.; Tack, B.F.  
J. Immunol. 136, 3407-3411, 1986  
A:Title: Structural analysis of human complement protein H: homology with C4b binding pr  
A:Reference number: S10479; MUID:86169701  
A:Accession: S10479  
A:Molecule type: mRNA  
A:Residues: 226-401, 'Y', 403-449 <KRI>  
A:Cross-references: GB:M1283; NID:g180472; PIDN:AAA52013.1; PID:g180473  
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner. Ho  
C:Comment: Alternative transcripts of 4.3, 1.8, and 1.4 kilobases are expressed in liver.  
C:Genetics: <HF1>  
A:Gene: GDB:HF1; HF  
A:Cross-references: GDB:120041; OMIM:134370  
A:Map position: 1q32-1q32  
C:Genetics: <HF2>  
A:Gene: GDB:HF2; HF  
A:Cross-references: GDB:129095  
A:Map position: 1q32-1q32  
A:Note: the correspondence between the two loci and the sequences indicated is unclear;  
C:Function:  
A:Description: a cofactor in the inactivation of C3b by serine proteinase I; also increa  
he alternative complement alternate pathway  
A:Pathway: complement alternate pathway  
C:Superfamily: complement factor H; complement factor H repeat homology  
C:Keywords: alternative splicing; complement alternate pathway; glycoprotein; plasma  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-449/Product: complement factor H, short splice form #status experimental <MAT>  
F:21-80/Domain: complement factor H repeat homology <FH01>  
F:85-141/Domain: complement factor H repeat homology <FH02>  
F:146-205/Domain: complement factor H repeat homology <FH03>  
F:210-262/Domain: complement factor H repeat homology <FH04>  
F:246-248/Region: cell attachment (R-G-D) motif  
F:267-320/Domain: complement factor H repeat homology <FH05>  
F:325-385/Domain: complement factor H repeat homology <FH06>  
F:389-442/Domain: complement factor H repeat homology <FH07>  
F:21-66,52-80,85-129,114-141,146-192,178-205,210-251,237-262,267-309,294-320,325-374,357  
F:217/Binding site: carbohydrate (Asn) (covalent) #status absent

Query Match 100.0%; Score 1497; DB 1; Length 449;  
Best Local Similarity 100.0%; Pred. No. 2,5e-106;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDCELPRLPRRTTELITGWSWSDQYPEGTOALYKCRPGYRSIGNYIMCRKGEWALNPLR 60  
|||  
Db 19 EDCELPRLPRRTTELITGWSWSDQYPEGTOALYKCRPGYRSIGNYIMCRKGEWALNPLR 78  
|||

QY 61 KCKRRPGHPRDPFGFTLTGWNVEYGVKAVYTCNGVOLLGEINTECDTGGWINDI 120  
|||  
Db 79 KCKRRPGHPRDPFGFTLTGWNVEYGVKAVYTCNGVOLLGEINTECDTGGWINDI 138  
|||

QY 121 PICGVVVCILPTAPENKIVSSAMEPREYHFGQAVRVCVSGKITEDEMHSCSDGFW 180  
|||  
Db 139 PICGVVVCILPTAPENKIVSSAMEPREYHFGQAVRVCVSGKITEDEMHSCSDGFW 198  
|||

QY 181 SKRPKCVELISCKSPDYINGSPISOKITLYKNEERFOYKCNNGYEYSEKGDVCTESGWRP 240  
|||  
Db 199 SKRPKCVELISCKSPDYINGSPISOKITLYKNEERFOYKCNNGYEYSEKGDVCTESGWRP 258  
|||

QY 241 LPSCSEKSCDNPYPYNGDYSPLRK 265  
|||  
Db 259 LPSCSEKSCDNPYPYNGDYSPLRK 283  
|||

RESULT 2  
NBHJH  
complement factor H precursor, long splice form - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text, change 22-Jun-1999  
C:Accession: S00254; A60238; A54726; A61565; A26505; I72654; S66298

R:Ripoche, J.; Day, A.J.; Harris, T.J.R.; Sim, R.B.  
Biochem. J. 249, 593-602, 1988  
A:Title: The complete amino acid sequence of human complement factor H.  
A:Reference number: S00254; MUID:88134059  
A:Accession: S00254  
A:Molecule type: mRNA  
A:Residues: 1-1231 <RIP>  
A:Cross-references: EMBL:Y00716; NID:g31964; PIDN:CAA68704.1; PID:g31965  
A:Note: 402-Tyr was also found  
R:Estaller, C.; Schaeble, W.; Dierich, M.; Weiss, E.H.  
Eur. J. Immunol. 21, 799-802, 1991  
A:Title: Human complement factor H: two factor H proteins are derived from alternat  
A:Reference number: A60238; MUID:91184292  
A:Accession: A60238  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 156,1177-1231 <EST>  
A:Note: only portions of this 4.3 kilobase mRNA were sequenced  
R:Day, A.J.; Ripoche, J.; Lyons, A.; McIntosh, B.; Harris, T.J.R.; Sim, R.B.  
Biosci. Rep. 7, 201-207, 1987  
A:Title: Sequence analysis of a cDNA clone encoding the C-terminal end of human compl  
A:Reference number: A54726; MUID:88025472  
A:Accession: A54726  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 579-1231 <DAY>  
A:Cross-references: GB:M17517; NID:g180497; PIDN:AAA52016.1; PID:g180498  
A:Note: parts of this sequence were determined by protein sequencing  
R:Ripoche, J.; Day, A.J.; Willis, A.C.; Belt, K.T.; Campbell, R.D.; Sim, R.B.  
Biosci. Rep. 6, 65-72, 1986  
A:Title: Partial characterization of a human complement factor H by protein and cDNA se  
A:Reference number: A61565; MUID:8618123  
A:Accession: A61565  
A>Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1050-1057, 'T', 1059-1102 <R12>  
R:Sim, R.B.; Discipio, R.G.  
Biochem. J. 205, 285-293, 1982  
A:Title: Purification and structural studies on the complement-system control protein  
A:Reference number: A26505; MUID:83048213  
A:Accession: A26505  
A:Molecule type: protein  
A:Residues: 19-20, 'Q', 22-29, 'V', 31-33, 'Q', 35 <SIM>  
R:Barlow, P.N.; Norman, D.G.; Steinkasserer, A.; Horne, T.J.; Pearce, J.; Discipio, P  
Biochemistry 31, 3626-3634, 1992  
A:Title: Solution structure of the fifth repeat of factor H: A second example of the  
A:Reference number: A44551; MUID:92232649  
A:Contents: annotation: NMR structure determination, residues 264-292  
R:Norman, D.G.; Barlow, P.N.; Baron, M.; Day, A.J.; Sim, R.B.; Campbell, I.D.  
J. Mol. Biol. 219, 717-725, 1991  
A:Title: Three-dimensional structure of a complement control protein module in soluti  
A:Reference number: A49224; MUID:91278097  
A:Contents: annotation: NMR structure determination, residues 927-985  
R:Estaller, C.; Koistinen, V.; Schaeble, W.; Dierich, M.P.; Weiss, E.H.  
J. Immunol. 146, 3190-3196, 1991  
A:Title: Cloning of the 1.4-kb mRNA species of human complement factor H reveals a no

A:Reference number: I56100; MUID:91201892  
A:Accession: I72654  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1047-1231 <RES>  
A:Cross-references: GB:M65294; NID:g183766; PIDN:AAA35948.1; PID:g183767  
R:Carroll, J.A.; Bates, R.C.; Smith, A.I.; Tetoz, T.; Arellano, A.; Gordon, D.L.; Burn  
Biochim. Biophys. Acta 1289, 305-311, 1996  
A:Title: Factor H co-purifies with thrombospondin isolated from platelet secretate.  
A:Reference number: S66298; MUID:96205365  
A:Accession: S66298  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 411-419,574-578,580-582 <CAR>  
C:Comment: Factor H has also been found bound to cell membranes in an unknown manner.

Query Match	100.0%	Score 1497	DB 1	Length 1231
Best Local Similarity	100.0%	Pred. NO. 7.5e-106		
Matches 265	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	EDCNELPRRNTETILTGSMDOQTYPEGTQAIYKCPKPYRSLGNVIMVORKGEMVNLPLR	60		
Db 19	EDCNELPRRNTETILTGSMDOQTYPEGTQAIYKCPKPYRSLGNVIMVORKGEMVNLPLR	78		
QY 61	KQKRRPCGHPDTPRGCTTTLTGNGNFEYGVKAVYTCNNGCYQLLGGTINRRECDTGWINDI	120		
Db 79	KQKRRPCGHPDTPRGCTTTLTGNGNFEYGVKAVYTCNNGCYQLLGGTINRRECDTGWINDI	138		
QY 121	PICEVVKCLPTVAPANGKIVTSSAMPEPREYHFGQAVRVCVNSGYITIEDDEMHGSDDOGF	180		
Db 139	PICEVVKCLPTVAPANGKIVTSSAMPEPREYHFGQAVRVCVNSGYITIEDDEMHGSDDOGF	198		
QY 181	SKERPKCVIEISCKSPDVINGSPISOKIITYKENERFOYKCNMGVEYSERGDVACTESGWRP	240		
Db 199	SKERPKCVIEISCKSPDVINGSPISOKIITYKENERFOYKCNMGVEYSERGDVACTESGWRP	258		
QY 241	LPSCEKSCDNPYPINPGDYSPLRK	265		
Db 259	LPSCEKSCDNPYPINPGDYSPLRK	283		

NEMSH  
 complement factor H precursor - mouse  
 N:Alternate names: protein beta-1-H  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text\_change 22-Jun-1999  
 C:Accession: A26154; 149711; 149728  
 R:Kristensen, T.; Tack, B.F.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 3963-3967, 1986  
 A:Title: Murine protein H is comprised of 20 repeating units, 61 amino acids in length  
 A:Reference number: A26154; MUID:86233353  
 A:Accession: A26154  
 A:Molecule type: mRNA  
 A:Residues: 1-1234 <RRI>  
 A:Cross-references: GB:M12660; NID:g193724; PIDN:AAA37759.1; PID:g367181  
 R:Natsunume-Sakai, S.; Nonaka, M.; Nonaka, M.; Harada, Y.  
 J. Immunol. 144, 358-362, 1990  
 A:Title: Demonstration of an unusual allelic variation of mouse factor H by the comp  
 A:Reference number: 149711; MUID:90111033  
 A:Accession: 149711  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-18 <RES>  
 A:Cross-references: GB:M31979; NID:g193726; PIDN:AAA37762.1; PID:g193729  
 R:Munoz-Canoves, P.; Tack, B.F.; Vilk, D.P.  
 Biochemistry 28, 9891-9897, 1989  
 A:Title: Analysis of complement factor H mRNA expression: Dexmethasone and IFN-gamma  
 A:Reference number: 149728; MUID:90146935  
 A:Accession: 149728  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-19 <RE2>  
 A:Cross-references: GB:J02891; NID:g193805; PIDN:AAA37795.1; PID:g553926  
 C:Comment: Two codominant alleles of factor H are present in mice.  
 C:Comment: Factor H functions as a cofactor in the inactivation of C3b by serine prote  
 (CS convertase) in the alternative complement pathway.  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: complement factor H; complement factor H repeat homology  
 C:Keywords: complement alternate pathway: duplication; glycoprotein; plasma  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-1234/Product: complement factor H #status predicted <MPT>  
 F:21-80/Domain: complement factor H repeat homology <FH01>  
 F:85-141/Domain: complement factor H repeat homology <FH02>  
 F:146-205/Domain: complement factor H repeat homology <FH03>  
 F:210-262/Domain: complement factor H repeat homology <FH04>  
 F:246-248/Region: cell attachment (R-G-D) motif  
 F:267-320/Domain: complement factor H repeat homology <FH05>  
 F:325-385/Domain: complement factor H repeat homology <FH06>  
 F:389-442/Domain: complement factor H repeat homology <FH07>  
 F:448-505/Domain: complement factor H repeat homology <FH08>  
 F:509-564/Domain: complement factor H repeat homology <FH09>  
 F:569-622/Domain: complement factor H repeat homology <FH10>  
 F:629-683/Domain: complement factor H repeat homology <FH11>  
 F:690-743/Domain: complement factor H repeat homology <FH12>  
 F:752-802/Domain: complement factor H repeat homology <FH13>  
 F:806-861/Domain: complement factor H repeat homology <FH14>  
 F:867-931/Domain: complement factor H repeat homology <FH15>  
 F:936-989/Domain: complement factor H repeat homology <FH16>  
 F:994-1048/Domain: complement factor H repeat homology <FH17>  
 F:1053-1107/Domain: complement factor H repeat homology <FH18>  
 F:1114-1168/Domain: complement factor H repeat homology <FH19>  
 F:1172-1233/Domain: complement factor H repeat homology <FH20>  
 F:121-66,552-129,114-144,146-192,178-205,210-251,237-262,267-309,294-320,325-374,374-408,408-580,836-861,867-920,906-931,936-978,964-989,994-1037,1023-1048,1053-1096,1096-1233,1233-1266,1266-1300,1300-1330,1330-1360,1360-1390,1390-1420,1420-1450,1450-1480,1480-1510,1510-1540,1540-1570,1570-1600,1600-1630,1630-1660,1660-1690,1690-1720,1720-1750,1750-1780,1780-1810,1810-1840,1840-1870,1870-1900,1900-1930,1930-1960,1960-1990,1990-2020,2020-2050,2050-2080,2080-2110,2110-2140,2140-2170,2170-2200,2200-2230,2230-2260,2260-2290,2290-2320,2320-2350,2350-2380,2380-2410,2410-2440,2440-2470,2470-2500,2500-2530,2530-2560,2560-2590,2590-2620,2620-2650,2650-2680,2680-2710,2710-2740,2740-2770,2770-2800,2800-2830,2830-2860,2860-2890,2890-2920,2920-2950,2950-2980,2980-3010,3010-3040,3040-3070,3070-3100,3100-3130,3130-3160,3160-3190,3190-3220,3220-3250,3250-3280,3280-3310,3310-3340,3340-3370,3370-3400,3400-3430,3430-3460,3460-3490,3490-3520,3520-3550,3550-3580,3580-3610,3610-3640,3640-3670,3670-3700,3700-3730,3730-3760,3760-3790,3790-3820,3820-3850,3850-3880,3880-3910,3910-3940,3940-3970,3970-4000,4000-4030,4030-4060,4060-4090,4090-4120,4120-4150,4150-4180,4180-4210,4210-4240,4240-4270,4270-4300,4300-4330,4330-4360,4360-4390,4390-4420,4420-4450,4450-4480,4480-4510,4510-4540,4540-4570,4570-4600,4600-4630,4630-4660,4660-4690,4690-4720,4720-4750,4750-4780,4780-4810,4810-4840,4840-4870,4870-4900,4900-4930,4930-4960,4960-4990,4990-5020,5020-5050,5050-5080,5080-5110,5110-5140,5140-5170,5170-5200,5200-5230,5230-5260,5260-5290,5290-5320,5320-5350,5350-5380,5380-5410,5410-5440,5440-5470,5470-5500,5500-5530,5530-5560,5560-5590,5590-5620,5620-5650,5650-5680,5680-5710,5710-5740,5740-5770,5770-5800,5800-5830,5830-5860,5860-5890,5890-5920,5920-5950,5950-5980,5980-6010,6010-6040,6040-6070,6070-6100,6100-6130,6130-6160,6160-6190,6190-6220,6220-6250,6250-6280,6280-6310,6310-6340,6340-6370,6370-6400,6400-6430,6430-6460,6460-6490,6490-6520,6520-6550,6550-6580,6580-6610,6610-6640,6640-6670,6670-6700,6700-6730,6730-6760,6760-6790,6790-6820,6820-6850,6850-6880,6880-6910,6910-6940,6940-6970,6970-7000,7000-7030,7030-7060,7060-7090,7090-7120,7120-7150,7150-7180,7180-7210,7210-7240,7240-7270,7270-7300,7300-7330,7330-7360,7360-7390,7390-7420,7420-7450,7450-7480,7480-7510,7510-7540,7540-7570,7570-7600,7600-7630,7630-7660,7660-7690,7690-7720,7720-7750,7750-7780,7780-7810,7810-7840,7840-7870,7870-7900,7900-7930,7930-7960,7960-7990,7990-8020,8020-8050,8050-8080,8080-8110,8110-8140,8140-8170,8170-8200,8



```

OY      8 PRRNRIE-----LTGSMDSQTEPTQAIYKCRGYSLSLGVN--MPCRKKEWALNPL 59
Db      24 PRRNRYVSLRAYVNITN--SSGSYPNGITLLQYTCRKGY--IGQIDQITVCVMNMVVPN-- 77
OY      60 RKQCRPGCHGPDPEFGFTLLTGNVPEYGVKAYVTCNEGYOLLGEINVRREC--DTDG-- 115
Db      78 -ECQRRRCSTPADLLNGWYTVV--GNLY--YGSVIYTCVMGYQLLSPT--SSGLGPDGRV 133
OY      116 -WTNDIPCEYVKKCLPVTAPENGKIVSSAMEPDRHY--FGQAVRFVCSNGYK--IEGDEE 171
Db      134 NWTPPRPICETIKCKRPPTLIANGTHTNI-----KEYVIYLDAAVYSCNDETKLTLTGSS 188
OY      172 MHCSDGDM--SKERKCAVEISCKSPDVINGSPISQKIITYKENERQY---KCNMKEYESE 227
Db      189 KQCSSTGKRWPDDEETKCEKVCCKIPQVANGHYEVRKT--SNWQYQVYNIKDKMGKGFRLQG 246
OY      228 RGDVAETESGWRP--LPSCEEKS---CDNPYIIPNGD 258
Db      247 ETPNMCKNGVMPFALPTCEKRAPRPGDMPHIDSGE 261

RESULT      7
NBH0C4
C4b-binding protein alpha chain precursor - human
N:Alternate names: C4BP; proline-rich protein
C:Species: Homo sapiens (man)
C:Date: 13-Aug-1986 #sequence, revision 30-Jun-1993 #text, change 22-Jun-1999
C:Accession: A33568; S02372; A90326; A24182; A93134; S29492; A31785; I52244; A03210
R:Matsumuchi, T.; Okamura, S.; Asou, T.; Sata, T.; Nihou, Y.
Biochem. Biophys. Res. Commun. 165, 138-144, 1989
A:Title: Molecular cloning of the cDNA coding for proline-rich protein (PRP): identity c
A:Reference number: A33568; MUID:90073699
A:Accession: A33568
A:Molecule type: mRNA
A:Residues: 1-597 <MA1>
A:Cross-references: GB:M1452; NID:g190501; PIDN:AA36507.1; PID:g190502
A>Note: the authors translated the codon GGA for residue 492 as Glu
R:Lintin, S.J.; Lewin, A.R.; Reid, K.B.M.
FEBS Lett. 232, 328-332, 1988
A:Title: Derivation of the sequence of the signal peptide in human C4b-binding protein
A:Reference number: S02372; MUID:88242821
A:Accession: S02372
A:Molecule type: mRNA
A:Residues: 17-81 <LI2>
A:Cross-references: EMBL:X07853
A>Note: although the sequence determined extends to residue 9 above, these authors assig
Biochem. J. 230, 133-141, 1985
A:Title: Molecular cloning and characterization of the cDNA coding for C4b-binding prote
A:Reference number: A90326; MUID:86025405
A:Accession: A90326
A:Molecule type: mRNA
A:Residues: 80-597 <CH2>
A:Cross-references: GB:X02865; NID:g29564; PIDN:CAA26617.1; PID:g29565
A>Note: 92-Thr and 357-His were also found
R:Lintin, S.J.; Reid, K.B.M.
FEBS Lett. 204, 77-81, 1986
A:Title: Studies on the structure of the human C4b-binding protein gene.
A:Reference number: A24182; MUID:86301119
A:Accession: A24182
A:Molecule type: DNA
A:Residues: 203-288 <LIN>
A:Cross-references: EMBL:X04284; EMBL:X04296
R:Rodríguez de Cordoba, S.; Sanchez-Corral, P.; Rey-Campos, J.
J. Exp. Med. 173, 1073-1082, 1991
A:Title: Structure of the gene coding for the alpha polypeptide chain of the human comp
A:Reference number: A43023; MUID:91217619
A:Contents: annotation; exon-intron boundaries
R:Chung, L.P.; Gagnon, J.; Reid, K.B.M.
Mol. Immunol. 22, 427-435, 1985
A:Title: Amino acid sequence studies of human C4b-binding protein: N-terminal sequence
A:Reference number: A93134; MUID:85296001

```

```

A:Accession: A03134
A:Molecule type: protein
A:Residues: 49-81 <CHI>
A>Note: This paper reports amino-terminal sequences of the intact protein and of a nu
R:Hessing, M.; Kanfers, D.; Takeya, H.; van't Veer, C.; Hackney, T.M.; Iwanaga, S.; E
FEBS Lett. 317, 228-232, 1993
A>Title: The region Ser(333)-Arg(336) of the alpha-chain of human C4b-binding protein
A:Reference number: S29492; MUID:93146164
A:Accession: S29492
A>Status: preliminary
A:Molecule type: protein
A:Residues: 381-404 <HES>
R:Suzuki, K.; Nishioke, J.
J. Biol. Chem. 263, 17034-17039, 1988
A>Title: Binding site for vitamin K-dependent protein S on complement C4b-binding pro
A:Reference number: A31785; MUID:89034204
A:Accession: A31785
A:Molecule type: protein
A:Residues: 495-505, 'X', 507-510, 'X', 512-515 <SUZ>
A>Note: This peptide appears to bind protein S
R:Idnaback, B.; Smith, C.A.; Muller-Eberhard, H.J.
Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465, 1983
A>Title: Visualization of human C4b-binding protein and its complexes with vitamin K
A:Reference number: A93950; MUID:83221615
A:Contents: annotation; electron microscopy; three-dimensional structure; ligand bindin
R:Aso, T.; Okamura, S.; Matsuguchi, T.; Sakamoto, N.; Sata, T.; Niho, Y.
Biochem. Biophys. Res. Commun. 174, 222-227, 1991
A>Title: Genomic organization of the alpha chain of the human C4b-binding protein gen
A:Reference number: I52244; MUID:91113199
A:Accession: I52244
A>Status: translation not shown; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-597 <ASO>
A:Cross-references: GB:M62486; NID:g190498; PIDD:AAA36506.1; PID:g190500
C:Comment: C4BP controls the classical pathway of complement activation. It binds as
the degradation of the C4bC2a complex (C3 convertase) by dissociating the complement
C:Comment: C4BP occurs in plasma in two forms, both of which bind complement fragment
insultide bonds. The beta chain binds the vitamin K-dependent plasma protein S. A mino
actors V and VIII.
C:Comment: The molecule has a central body supporting seven tentacles (alpha chains),
C:Genetics:
A:Gene: GDB:C4BPA
A:Cross-references: GDB:120568; OMTM:120830
A:Map position: 1q32-1q32
A:Introns: 48/1; 110/1; 143/2; 172/1; 236/1; 297/1; 362/1; 425/1; 482/1; 540/3
C:Complex: octamer of seven alpha chains and one beta chain
C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology
C:KeyWords: acute phase; chylomycin; complement factor H repeat homology
F:148/Domain: signal sequence #status predicted <SIG>
F:49-597/Product: C4b-binding protein alpha chain #status predicted <MAT>
F:50-108/Domain: complement factor H repeat homology <FH1>
F:113-110/Domain: complement factor H repeat homology <FH2>
F:175-234/Domain: complement factor H repeat homology <FH3>
F:233-234/Domain: complement factor H repeat homology <FH4>
F:239-360/Domain: complement factor H repeat homology <FH5>
F:366-422/Domain: complement factor H repeat homology <FH6>
F:381-404/Region: complement C4b binding #status predicted
F:425-480/Domain: complement factor H repeat homology <FH7>
F:484-538/Domain: complement factor H repeat homology <FH8>
F:521, 506, 528/Binding site: carbohydrate (Asn) (covalent) #status experimental

```



QY 129 LPYAPENGKIVSSAMEPDBREYHGOAVRFVNSGKYLEGDEMHCSDD---GFWSEK 184  
 176 KPPDPDINGHSGE---ENRYAAGFSVTYSCDPFLLGHSAGISYVENTITGVMPSP 231  
 QY 185 PKVEISCKSPDYINGSPISO-KIITYENRFOYKCMNGEYISRGDAV--C-TESGMRP 240  
 232 PTECKITCRPRPDVSHGEMVSGFPIYMKDITVFCOKGFVL--RGSSVLIHCADSKWNP 289  
 QY 241 L-PSCEKSCDN--PIYINGDY 259  
 290 SPSPACEPNSCINLPDIFHAW 310

RESULT 8  
 S53711  
 C:ABP alpha chain precursor - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)  
 C:Date: 10-Sep-1999 #sequence, revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S53711  
 R:de Frutos, P.G.; Dahlbaeck, B.  
 Biochim. Biophys. Acta 1261, 285-289, 1995  
 A:Title: CDNA structure of rabbit C4b-binding protein alpha-chain. Preserved sequence and  
 A:Reference number: S53711, MUID:95226458  
 A:Accession: S53711  
 A:Status: preliminary; nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-597 <DEF>  
 A:Cross-references: EMBL:235490  
 C:Superfamily: C4b-binding protein alpha chain; complement factor H repeat homology  
 F:50-107/Domain: complement factor H repeat homology <FH1>  
 F:112-169/Domain: complement factor H repeat homology <FH2>  
 F:174-234/Domain: complement factor H repeat homology <FH3>  
 F:239-294/Domain: complement factor H repeat homology <FH4>  
 F:299-360/Domain: complement factor H repeat homology <FH5>  
 F:364-422/Domain: complement factor H repeat homology <FH6>  
 F:426-480/Domain: complement factor H repeat homology <FH7>  
 F:484-538/Domain: complement factor H repeat homology <FH8>

Query Match 20.2%; Score 302; DB 1; Length 597;  
 Best Local Similarity 32.0%; Pred. No. 1.4e-15;  
 Matches 87; Conservative 37; Mismatches 118; Indels 30; Gaps 16;  
 QY 2 DCNELPRRRTTEILTSGMSQTYPEGTQAIYKCRPKYRSLG-NVIWCR-KGEVVALNPL 59  
 49 DCGP-PHLLFASISISENEYOTGITLKYTCRPGYTRNLNPLILCKCPGLM---SYD 104  
 QY 60 RKCKRCRGGHGPDPFTFTLTGNGVEYGVKAVYTCNEGVLGELTYRECDPD---G 115  
 105 TFCYKRCRNPGLPNOVEYK--TDFSFGQIFSCSEGITILIGSTT-SHCDIQEKGYE 161  
 QY 116 WTNDIPICEVVKCLPYTAPENGKIVSSAMEPDBREYH-FGOAVRFVNSGKYLEGDEMH 174  
 162 WSDPLPCEIYKCEPPNIINGKHNG---NEDIHTYGSVTSYSCNDRFSLGCASTISC 217  
 QY 175 S---DDGFWSEKPKVEISCKSPDYINGSPISO-KIITYENRFOYKCMNGEYISRG 229  
 218 TVNKNIVGVSPSPVCKETITCSPNPVPHGKITISFGPIYMKDSIMWTCIDGFLV--RG 275  
 QY 230 DAV--C-TESGMRPLPS-CEKSKC-DNPIYIPN 256  
 276 SSLIHCELDCKMNPSPVCESSNCLGLPNVPH 307

RESULT 9  
 MWVZSP  
 A:apolipoprotein H homolog precursor - vaccinia virus  
 N:Alternate names: 35K secretory protein, C3L protein, virokin  
 C:Species: vaccinia virus  
 A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1989 #sequence, revision 30-Jun-1990 #text\_change 22-Jun-1999  
 C:Accession: A31005, B42504

R:Kotwal, G.J.; Moss, B.  
 Nature 335, 176-178, 1988  
 A:Title: Vaccinia virus encodes a secretory polypeptide structurally related to compl  
 A:Reference number: A31005, MUID:88318974

A:Accession: A31005  
 A:Molecule type: DNA  
 A:Residues: 1-263 <KOT>  
 A:Cross-references: GB:X13166; NID:g60690; PIDN:CAA31564.1; PID:g60691  
 A:Experimental source: strain WR  
 R:Goebel, S.J.; Johnson, G.P.; Petkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.  
 Virology 179, 517-563, 1990  
 A:Title: Appendix to "The complete DNA sequence of vaccinia virus".  
 A:Reference number: A42501  
 A:Accession: B42504  
 A:Molecule type: DNA  
 A:Residues: 1-263 <GOE>  
 A:Cross-references: GB:M35027; NID:g335317; PIDN:AAA47997.1; PID:g335345  
 A:Experimental source: strain Copenhagen  
 R:Goebel, S.J.; Johnson, G.P.; Petkus, M.E.; Davis, S.W.; Winslow, J.P.; Paoletti, E.  
 Virology 179, 247-266, 1990  
 A:Title: The complete DNA sequence of vaccinia virus.  
 A:Reference number: A42531; MUID:91021027  
 A:Contents: annotation; possible protein-coding frames  
 A:Note: neither amino acid nor nucleotide sequence is given  
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom  
 C:Keywords: duplication; extracellular protein  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-263/Product: C4b-binding protein homolog #status predicted <MAT>  
 F:21-81/Domain: complement factor H repeat homology <FH1>  
 F:86-143/Domain: complement factor H repeat homology <FH2>  
 F:148-201/Domain: complement factor H repeat homology <FH3>  
 F:206-261/Domain: complement factor H repeat homology <FH4>

Query Match 20.1%; Score 301.5; DB 1; Length 263;  
 Best Local Similarity 33.3%; Pred. No. 6.1e-16;  
 Matches 78; Conservative 29; Mismatches 106; Indels 21; Gaps 12;  
 QY 20 SDQTYPEGTQAIYKCRPKYR--SLGNVIWCRKGEVVALNPLRCKCRPGHDPDPGCT 77  
 40 ANANYNIGDTLEYLCLPGRKOKMGPYAKCTGTGWLTFN--OCIRKRCSPRDINDNQ 96  
 QY 78 FTLLGNGVVEYGVKAVYTCNEGVLGELTYRECDPDG---WTNDIPICEVVKCLPYTA 133  
 97 LDIGG---VDRGSSITTSYSCNGYHLIGESKSYCELGSGSWMWPEAPICSVACQSPS 153  
 QY 134 PENCKIVSSAMEPDBREYHGOAVRFVNSGKYLEGDEMHCSDDGFWSEKPKVEISCK 193  
 154 ISNR--HNGEYD--FTDGSVTVYSCNSGYSLGNSCVLCS-GGEMS-DPPTCQIYKCP 207  
 QY 194 SPDYINGSPISO-KIITYENRFOYKCMNGEYISRGDAVTESG-WRP-LPSC 244  
 208 HPTISNGLSGFRFSYSDNDVDFCKKYGYKILSGSSSTCSPGNTWPKELPKC 261

RESULT 10  
 A45900  
 Complement C3b receptor type 2 long form precursor - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1993 #sequence, revision 30-Sep-1993 #text\_change 08-Oct-1999  
 C:Accession: A45900; I48306  
 R:Kurtz, C.B.; O'Toole, E.; Christensen, S.M.; Wells, J.H.  
 J. Immunol. 144, 3581-3591, 1990  
 A:Title: The murine complement receptor gene family. IV. Alternative splicing of C12  
 A:Reference number: A45900; MUID:90229734  
 A:Accession: A45900  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-676 <KUR>  
 A:Cross-references: GB:M36470  
 A:Experimental source: clone 31-1  
 R:Kilm, Y.U.; Kinoshita, T.; Molina, H.; Hourcade, D.; Seya, T.; Wagner, L.M.; Holers,  
 J. Exp. Med. 181, 151-159, 1995









